

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Análise longitudinal da mudança terapêutica nos doentes seropositivos para o VIH

Andreia Heitor Martins da Cunha Leite

TRABALHO DE PROJETO

MESTRADO EM BIOESTATÍSTICA

2012

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Análise longitudinal da mudança terapêutica nos doentes seropositivos para o VIH

Andreia Heitor Martins da Cunha Leite

TRABALHO DE PROJETO ORIENTADO PELA
PROF.^a DOUTORA VALESKA ANDREOZZI

MESTRADO EM BIOESTATÍSTICA

2012

Resumo

Introdução: A terapêutica antiretroviral (TAR) diversificou-se nos últimos anos, sendo complexa a decisão da mudança terapêutica. Alguns estudos internacionais descrevem os fatores associados à mudança terapêutica não sendo consensuais nos resultados obtidos.

Objetivos: Descrever os fatores associados à mudança terapêutica nos doentes seropositivos para o VIH-1, utilizando os modelos de análise de sobrevivência mais adequados às características do estudo.

Métodos: Foi realizado um estudo de coorte retrospectiva com os indivíduos seropositivos para o VIH-1, seguidos no Hospital de Santa Maria. A informação constante no processo clínico foi registada para uma amostra aleatória de 320 indivíduos com pelo menos duas dispensas de terapêutica entre 2005 e 2008. Foram consideradas mudanças as alterações em pelo menos um dos fármacos do regime. Os indivíduos foram censurados caso interrompessem a terapêutica. Foi utilizado o modelo de regressão de Cox para descrever os fatores associados à primeira mudança terapêutica e o modelo PWP para os fatores associados às mudanças até à quarta. Nos modelos construídos a seleção das covariáveis candidatas foi feita com recurso ao método *stepwise* considerando um valor-p de inclusão de 0,20 e de exclusão de 0,25.

Resultados: Da amostra inicial (n=320) 194 indivíduos foram incluídos. O regime inicial foi associado com a primeira mudança Razão de Riscos (RR): 1,777 [Intervalo de Confiança a 95% (IC95): 1,121 – 2,816]. A contagem de linfócitos CD4⁺ associou-se de forma estatisticamente significativa às mudanças terapêuticas até à quarta (RR: 0,999; IC95%: 0,998 – 0,9997).

Discussão: Verificou-se que os fatores associados à primeira mudança terapêutica diferem dos associados às mudanças terapêuticas até à quarta mudança. Embora tenha um efeito relativamente pequeno os doentes com contagens mais baixas de CD4⁺ devem ser acompanhados de perto visto que se encontram em maior risco de mudança terapêutica e por conseguinte de esgotar as opções disponíveis.

Palavras-chave: Infecção VIH/SIDA, Modelos de eventos múltiplos, Mudança terapêutica, Regressão de Cox.

Abstract

Introduction: Antiretroviral therapy (ART) has evolved across the last years being difficult to decide whether change or not a regimen. Even though some international studies had described factors associated with therapy switching the results are not consensual.

Aims: Describe factors associated with ART switching, using survival models adequate do the study characteristics.

Methods: A retrospective cohort study was conducted with HIV-1 patients followed at Hospital de Santa Maria. Data of clinical records was collected for a random sample of 320 patients with at least two pharmacy refills between 2005 and 2008. Switches were considered if individuals changed at least one drug. Subjects were censored if they stopped therapy. Cox regression was used to describe factors associated with first therapy switch and PWP model to describe factors associated with switches until the fourth. To fit models covariates were selected using stepwise method considering an inclusion p-value of 0.20 and an exclusion p-value of 0.25.

Results: From the initial sample (n=320) 194 subjects were included. Initial regimen was associated to the first switch – Hazard Ratio (HR) of protease inhibitors vs. non-nucleoside reverse transcriptase inhibitors: 1.777 [95% Confidence Interval (95%CI): 1.121 – 2.816]. CD4⁺ cells count was independently associated with switches until the fourth (HR: 0.999; 95%CI: 0.998 – 0.9997).

Discussion: Factors associated with first therapy switch differ from the ones associated with the switches until the fourth. Although a relatively small effect was observed patients with lower CD4⁺ counts should be carefully monitored as they have a higher risk of switching and exhausting therapy options.

Key-words: Cox Regression, HIV/AIDS infection, Multiple events models, Switching therapy.

Índice

Resumo	3
Abstract	4
Figuras	7
Tabelas	8
Abreviaturas	9
Introdução e Objetivos	11
<i>Frequência de mudança do regime</i>	<i>15</i>
<i>Razões de mudança</i>	<i>15</i>
<i>Fatores associados à mudança</i>	<i>16</i>
<i>Objetivo</i>	<i>23</i>
Metodologia	25
<i>O projeto</i>	<i>27</i>
<i>Variável resposta e evento de interesse</i>	<i>28</i>
<i>Regressão de Cox</i>	<i>28</i>
<i>Modelos de eventos múltiplos</i>	<i>30</i>
<i>Modelo Andersen-Gill:</i>	<i>30</i>
<i>Modelo Wei-Lin-Weissfeld (ou modelo marginal)</i>	<i>31</i>
<i>Modelo Prentice-Williams-Peterson (ou modelo condicional)</i>	<i>33</i>
<i>Resumo da utilização dos vários modelos</i>	<i>34</i>
<i>Criação da base de dados para análise</i>	<i>38</i>
<i>Covariáveis</i>	<i>41</i>
<i>Seleção de covariáveis</i>	<i>43</i>
<i>Construção do modelo – abordagem no software</i>	<i>44</i>
<i>Diagnóstico do modelo</i>	<i>44</i>
Resultados	47
<i>Características gerais da amostra em estudo</i>	<i>49</i>

<i>Modelo para o tempo até à 1ª mudança terapêutica</i>	52
<i>Análise Exploratória</i>	52
<i>Ajustamento do modelo</i>	54
<i>Qualidade de ajustamento</i>	55
<i>Diagnóstico do modelo</i>	55
<i>Modelo para múltiplas mudanças terapêuticas</i>	59
Discussão	61
<i>Modelo de Cox</i>	63
<i>Modelo para eventos múltiplos</i>	66
<i>Limitações do estudo</i>	67
<i>Alternativas à construção do modelo</i>	67
<i>Reflexão relativamente aos modelos de eventos múltiplos</i>	68
Conclusão	71
Agradecimentos	75
Bibliografia	77
Anexos	81
<i>Construção da base de dados a analisar</i>	81
<i>Análise exploratória</i>	91
<i>Seleção de covariáveis (Modelo de Cox)</i>	93
<i>Modelo Final (Modelo de Cox)</i>	95
Diagnóstico	96
<i>Resíduos de Schoenfeld (Figura 9)</i>	96
<i>Correlação linear dos resíduos das covariáveis com o tempo de sobrevivência...</i>	96
<i>Resíduos martingala e desvios residuais (Figura 10)</i>	96
<i>Resíduos score (Figura 11)</i>	97
<i>Seleção de covariáveis (Modelo PWP)</i>	97
<i>Modelo Final (Modelo PWP)</i>	98

Figuras

Figura 1 – Ciclo de replicação do vírus e alvos terapêuticos	14
Figura 2 – Representação esquemática das possíveis formulações de intervalos de risco – (a) <i>Gap time</i> , (b) Tempo total, (c) Processos de contagem.	34
Figura 3 – Representação esquemática dos riscos basais no modelo de Cox (A) e nos vários modelos de eventos múltiplos (B – Modelo Andersen-Gill – AG; C – Modelo Wei, Lin e Weissfeld – WLW; D - Modelo de Prentice, Williams e Peterson – PWP). 36	
Figura 4 – Representação gráfica do ano de diagnóstico	51
Figura 5 – Representação gráfica do tempo de seguimento, em meses	51
Figura 6 – Representação gráfica dos valores de linfócitos CD4 ⁺ no início da terapêutica HAART.....	51
Figura 7 – Representação gráfica dos valores de carga vírica de acordo com a transformação utilizada.....	52
Figura 8 – Curvas de sobrevivência de Kaplan-Meier para as variáveis em estudo	53
Figura 9 – Resíduos de Schoenfeld para o modelo final	56
Figura 10 – Representação gráfica dos resíduos martingala (A) e desvios residuais (B) contra o índice, desvios residuais contra os valores preditos (C) e gráfico quantil-quantil dos desvios residuais (D).....	57
Figura 11 – Resíduos <i>score</i> contra as variáveis incluídas..	58
Figura 12 – Risco acumulado para a ocorrência de cada mudança	60

Tabelas

Tabela 1 – Resumo dos estudos que avaliaram a mudança terapêutica	17
Tabela 2 – Variáveis incluídas no modelo final dos estudos considerados.....	20
Tabela 3 – Definição do intervalo de risco e grupo em risco para cada modelo.....	35
Tabela 4 – Equações das funções de risco e verossimilhanças parciais dos modelos apresentados.....	37
Tabela 5 – Características sociodemográficas, laboratoriais e de terapêutica dos 194 participantes da amostra estudada	49
Tabela 6 – Seleção de covariáveis pelo método <i>stepwise</i> (modelo de Cox)	54
Tabela 7 – Razões de riscos e respectivos intervalos de confiança a 95% de acordo com o modelo de Cox escolhido	54
Tabela 8 – Seleção de covariáveis candidatas pelo método <i>stepwise</i> (modelo PWP)....	59
Tabela 9 – Razões de riscos e erro padrão das variáveis incluídas no modelo ajustado pelo método PWP	59

Abreviaturas

ADN – Ácido desoxirribonucleico
ARN – Ácido ribonucleico
CV – Carga vírica
HAART – Terapêutica antiretroviral de alta eficácia
HSM – Hospital de Santa Maria
INTR – Inibidor nucleósido da transcriptase reversa
INNTR – Inibidor não-nucleósido da transcriptase reversa
IE – Inibidor de entrada
II – Inibidor da integrase
IP – Inibidor da protease
IP/r – Inibidor da protease potenciado por ritonavir
Modelo AG – Modelo de Andersen-Gill
Modelo PWP – Modelo de Prentice-Williams-Peterson
Modelo WLW – Modelo de Wei-Lin-Weissfeld
ND – Não detetável
SDI – Serviço de Doenças Infeciosas
SIDA – Síndrome da Imunodeficiência Adquirida
TAR – Terapêutica antiretroviral
VIH(-1) – Vírus da Imunodeficiência Humana (tipo 1)

Introdução e Objetivos

A infecção VIH/SIDA constitui um problema de saúde pública em Portugal, país que apresenta valores de incidência e prevalência superiores aos restantes países da Europa Ocidental [1].

Esta infecção tem na sua origem o Vírus da Imunodeficiência Humana (VIH), um vírus de ácido ribonucleico (ARN) da família *Lentivirus* que é responsável por uma infecção crónica, com um longo período de latência. Os linfócitos T CD4⁺ desempenham um papel central nesta infecção. O vírus infeta estas células e aproveita-se dos seus constituintes para se replicar e assim infetar novas células. Estes linfócitos infetados acabam por ser destruídos, sendo um dos mecanismos de destruição progressiva do sistema imune [2]. As manifestações clínicas relacionam-se diretamente com o vírus e, por outro lado, com a destruição progressiva do sistema imune que pode levar ao aparecimento de infeções oportunistas, neoplasias malignas e doenças autoimunes, anos após a infecção inicial [2, 3].

Desde a introdução da terapêutica antiretrovítica de alta eficácia (HAART) em 1996 que a mortalidade associada à infecção tem diminuído. A HAART caracteriza-se pela utilização, em simultâneo, de três ou mais fármacos com atividade contra o VIH (antiretrovíticos) [2]. Atualmente, existem 5 classes de fármacos antiretrovíticos (Figura 1) [2, 3]:

- **Inibidores nucleósidos/nucleótidos da transcriptase reversa (INTR)** – Inibem a síntese de ácido desoxirribonucleico (ADN) viral, ao serem incorporados como constituintes da cadeia de ADN. Ao impedirem a replicação vírica, evita-se a infecção de novos linfócitos.
- **Inibidores não-nucleósidos da transcriptase reversa (INNTR)** – Inibem diretamente a atividade da transcriptase reversa, a enzima responsável pela transformação do ARN viral em ADN que será integrado no ADN humano.
- **Inibidores da protease (IP)** – Inibem a enzima protease que é responsável pela maturação dos viriões após a saída dos linfócitos T CD4⁺.
- **Inibidores de entrada (IE)** – inibem a ligação do vírus aos recetores das células, evitando desta forma a sua entrada nas mesmas.
- **Inibidores da integrase (II)** – impedem a integração do DNA viral no DNA humano.

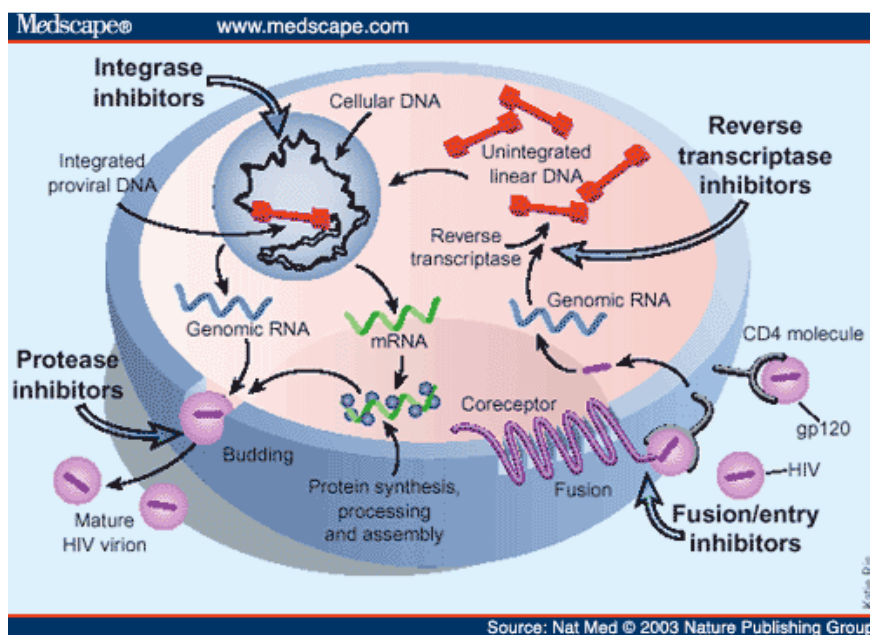


Figura 1 – Ciclo de replicação do vírus e alvos terapêuticos [4]

As recomendações para iniciar terapêutica antiretroviral modificaram-se ao longo dos anos. Atualmente o início está recomendado [5]:

- Nos doentes sintomáticos,
- Nos doentes assintomáticos com $CD4^+ < 350/mm^3$,
- Nos doentes assintomáticos com $CD4^+ > 350/mm^3$ caso as comorbilidades ou a evolução clínica assim o justifiquem.

O esquema inicial deve ser adequado ao doente mas recomenda-se que seja constituído por 3 fármacos – 2 INTRs (designados de “esqueleto”) e um de entre INNTR, IP/r (inibidor da protease potenciado por ritonavir) ou II [5]. Esta última classe teve aprovação mais recente (em 2008), assim como os IE (em 2003) [2]. Os IE são utilizados em doentes com múltiplos tratamentos prévios e resistências aos fármacos de primeira linha [2].

É importante escolher o tratamento certo para cada doente, dado que a terapêutica não é curativa e que o primeiro esquema de TAR é o que melhor garante o sucesso da mesma. Assim, a escolha do regime terapêutico e a decisão de mudar esse mesmo regime são decisões complexas que têm por objetivo preservar a supressão virológica e evitar a exaustão de opções terapêuticas futuras [5, 6].

Frequência de mudança do regime

Conhecer a frequência de mudança dos regimes terapêuticos e as razões a ele associadas na prática clínica reveste-se de particular importância. Este conhecimento permite-nos avaliar se o tratamento consegue ou não controlar a replicação viral ou se a questão se prende principalmente com questões de tolerância do tratamento, pelo doente [7].

Estudos internacionais avaliaram a mudança de terapêutica e reportaram que a incidência da primeira mudança varia de 25 a 65 por 100 pessoas/ano, sendo o tempo mediano até à mudança de 8 a 17,5 meses. Na Tabela 1 encontram-se resumidos os principais estudos internacionais bem como as observações chave deles resultantes.

López *et al.* acompanharam retrospectivamente uma coorte de 401 indivíduos durante 3 anos e 3 meses e descreveram o tempo até à quinta mudança. Nesta coorte, verificou-se que 401 indivíduos tiveram a sua primeira terapêutica alterada após um tempo mediano de 560 dias, em 234 indivíduos verificou-se uma segunda mudança após um tempo mediano de 360 dias, em 106 indivíduos houve uma terceira mudança após 330 dias (mediana) e em 34 indivíduos existiu uma quarta mudança após um tempo mediano de 202 dias [8].

A nível nacional são conhecidos dois resumos apresentados em congressos nacionais. Um deles descreve 137 doentes de Lisboa seguidos entre 1997 e 2004 com um tempo médio até à mudança de $14,0 \pm 4,1$ meses, e uma média de $2,2 \pm 1,7$ mudanças por doente [9]. O segundo estudo descreve a casuística da consulta de imunodeficiência dos Hospitais da Universidade de Coimbra, de 2007 a 2009, reportando um número médio de alterações do esquema terapêutica de 1,1. Contudo estes dados não nos permitem comparar com os resultados internacionais e não são do nosso conhecimento publicações portuguesas acerca desta questão [10].

Razões de mudança

As razões principais para mudar os regimes terapêuticos parecem relacionar-se com efeitos adversos, falência imunológica/viológica, comorbilidades e não-adesão, entre outros [10, 11]. Em estudos anteriores os efeitos adversos foram considerados a razão principal para as mudanças (cerca de 25% dos doentes que mudaram ou pararam a terapêutica), sendo a intolerância gastrointestinal o efeito adverso mais comum (44% dos que pararam terapêutica devido a uma reação adversa) [6, 12].

Apesar de vários trabalhos terem descrito este fenómeno, em geral, apenas a primeira mudança foi caracterizada. López *et al.*, que descreveram os motivos até à terceira mudança, relatam o predomínio dos efeitos adversos em todas as mudanças [8].

Fatores associados à mudança

Vários estudos internacionais avaliaram os fatores associados à mudança e/ou paragem terapêutica. Na Tabela 1 encontra-se um resumo dos estudos e principais conclusões, salientando-se a heterogeneidade ao nível metodológico, amostral e temporal. A nível estatístico a maioria dos estudos que caracterizam os fatores independentemente associados à mudança fazem-no através do modelo de riscos proporcionais de Cox [6, 12–24], considerando a variável resposta como sendo tempo até à mudança. A regressão logística, assumindo a variável resposta (mudança) como dicotómica, [19, 25–31] ou como uma variável nominal com vários níveis correspondentes ao tipo/motivo de mudança (regressão logística multinominal) [32–34] é o modelo que se segue. No geral, estes estudos analisam apenas a primeira mudança terapêutica. Robison *et al.* analisaram a mudança com base nos regimes terapêuticos (e não com base nas mudanças por doente) mas, admitindo a existência de mais do que um regime por pessoa, introduziram uma correção no modelo devido à provável existência de dependência (através dos estimadores robustos de variância de Huber-White) [33]. Para além dos estudos referidos, Srasuebkul *et al.* abordaram a questão de forma um pouco diferente modelando não o tempo até à mudança ou a mudança como variável dicotómica mas permitindo a possibilidade de várias mudanças por indivíduo e utilizando para tal um modelo de regressão de Poisson com efeitos aleatórios [35].

Tabela 1 – Resumo dos estudos que avaliaram a mudança terapêutica

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	PERÍODO ANALISADO	N	FOLLOW-UP (MESES)*	VARIÁVEL(EIS) RESPOSTA**	METODOLOGIA	COMENTÁRIOS
van Roon <i>et al.</i> , Holanda, 1998 [13]	1996-1998	99	15,0±0,3	Mudança/paragem	Modelo de riscos proporcionais de Cox	Apenas para regimes com IP
Monforte <i>et al.</i> , Itália, 2000 [14]	1997-1999	862	11,3	Mudança/paragem por motivo	Modelo de riscos proporcionais de Cox	A população em estudo foi restrita aos homens
Hänsel <i>et al.</i> , Suíça, 2001 [25]	1996-1999	236	-	Mudança por motivo e paragem	Regressão logística	-
Mocroft <i>et al.</i> , RU, 2001 [16]	1996-1998	556	14,2	Mudança e paragem	Modelo de riscos proporcionais de Cox	-
O'Brien <i>et al.</i> , EUA, 2003 [6]	1997-2001	345	8,1	Mudança/paragem em geral e por motivo	Modelo de riscos proporcionais de Cox	Apenas foram considerados os 6 esquemas terapêuticos mais frequentes no centro do estudo
Kirstein <i>et al.</i> , EUA, 2002 [7]	1994-2000	994	-	Mudança/paragem	Modelo de riscos proporcionais de Cox	A população em estudo foi restrita às mulheres
Dieleman <i>et al.</i> , Holanda, 2002 [17]	1996-1999	775	-	Mudança	Modelo de riscos proporcionais de Cox	Analisa a 2ª mudança por efeitos adversos
López <i>et al.</i> , Espanha, 2002 [8]	1997-2000	401	-	-	Descritivo	-
Dorruci <i>et al.</i> , Itália, 2002 [15]	1997-2000	2002	11 meses – regimes com IP 5-6 meses – regimes com INNTR	Mudança por motivo	Modelo de riscos proporcionais de Cox	-
Ripamonti <i>et al.</i> , Itália, 2004 [36]	1997-2000	465	-	Mudança/paragem	Log-binomial	-
Kerr <i>et al.</i> , Canadá, 2005 [26]	1996-2001	160	-	Paragem	Regressão logística	A população em estudo foi restrita aos IDU

Tabela 1 (cont.) – Resumo dos estudos que avaliaram a mudança terapêutica

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	PERÍODO ANALISADO	N	FOLLOW-UP (MESES)*	VARIÁVEL(EIS) RESPOSTA**	METODOLOGIA	COMENTÁRIOS
Li <i>et al.</i> , EUA, 2005 [19]	1997-2002	589 (2916 pares de visitas)	-	Mudança e paragem em separado	Regressão logística com controlo da dependência	A população em estudo foi restrita aos homens
Ahdieh-Grant <i>et al.</i> , EUA, 2005 [18]	1994-2001	936	-	Paragem	Modelo de riscos proporcionais de Cox	A população em estudo é restrita às mulheres
Yuan <i>et al.</i> , EUA, 2006 [20]	1996-2003	3414		Mudança por motivo	Modelo de riscos proporcionais de Cox	-
Kumarasamy <i>et al.</i> , Índia, 2006 [27]	1996-2004	1443	-	Mudança/paragem	Regressão logística	-
Kiguba <i>et al.</i> , Uganda, 2007[28]	2005-2006	686	-	Mudança e paragem em separado	Regressão logística	-
Srasuebkul <i>et al.</i> , Ásia, 2007[35]	2003-2009	1846	-	Mudança	Regressão de Poisson com efeitos aleatórios	Analisa até à 3ª mudança
Maggiolo <i>et al.</i> , Itália, 2007 [29]	1999-2006	582	1435 PAFU	Mudança/paragem	Regressão logística	Apenas regimes com NVP
de la Torre <i>et al.</i> , Espanha, 2008 [12]	1997-2003	603	21,8	Mudança	Modelo de riscos proporcionais de Cox	-
Vo <i>et al.</i> , Suíça, 2008 [34]	2000-2005	1866	-	Mudança e paragem em separado	Regressão logística multinomial	-
Golub <i>et al.</i> , EUA, 2008 [32]	1996-2005	1555	-	Mudança	Modelo de riscos proporcionais de Cox	Define mudança apenas se os fármacos forem de classes diferentes. A população em estudo é constituída apenas por mulheres

Tabela 1 (cont.) – Resumo dos estudos que avaliaram a mudança terapêutica

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	PERÍODO ANALISADO	N	FOLLOW-UP (MESES)*	VARIÁVEL(EIS) RESPOSTA**	METODOLOGIA	COMENTÁRIOS
Robison <i>et al.</i> , EUA, 2008 [33]	-	738 (1852 regimes)	-	Mudança/paragem no geral e por motivo	Regressão logística multinominal corrigida para dependência (estimadores robustos de variância Huber-White)	Análise feita por regimes e não por indivíduos
Mirjam-Colette <i>et al.</i> , EUA, 2009 [31]	1999-2004	631	-	Mudança/paragem	Regressão logística	-
Himelhoch <i>et al.</i> , EUA, 2009 [30]	2000-2005	4989	-	Paragem	Regressão logística com função de ligação log-log	-
Sabin <i>et al.</i> , RU, 2009 [21]	1998-2005	8708	-	Mudança exceto por falência	Modelo de riscos proporcionais de Cox	-
Cardoso <i>et al.</i> , Brasil, 2010 [22]	1996- 2008	670	22,0	Mudança/paragem em geral e por motivo (falência e efeitos adversos)	Modelo de riscos proporcionais de Cox	-
Cesar <i>et al.</i> , América Latina, 2010 [23]	1996-2007	5026	-	Mudança/paragem	Modelo de riscos proporcionais de Cox	Faz imputação para os dados omissos
Landier <i>et al.</i> , 2010, Mali [24]	-	865	15 meses (16,2/100 PAFU)	Mudança	Modelo de riscos proporcionais de Cox	-

*Salvo indicação contrária

**Mudança/paragem refere-se à modelação simultânea dos acontecimentos, enquanto mudança e paragem se refere à construção de dois modelos diferentes.

Legenda: EUA – Estado Unidos da América; IDU – utilizadores de drogas endovenosas; INNTR – Inibidor não nucleósido da transcriptase reversa; IP – Inibidor da Protease; NVP – Nevirapina; PAFU – Pessoas-ano de *follow-up*; RU – Reino Unido.

Os vários estudos identificam fatores associados à mudança terapêutica em diversos níveis: sociodemográficos, relacionados com a terapêutica e com a infeção. Considerando as características do presente trabalho (ver Métodos), a Tabela 2 apresenta os fatores associados à mudança terapêutica nos estudos em que o *outcome* foi precisamente a mudança terapêutica, sem serem restritos a apenas um determinado regime terapêutico ou fármaco [7, 12, 16, 19, 24, 34].

Os fatores identificados variam com o estudo e a metodologia. Contudo, alguns, pela frequência com que são referidos, merecem destaque: a carga vírica, cujo aumento, aparece geralmente associado a um maior risco de mudança [16, 19, 34], a contagem de células CD4⁺, com resultados variáveis nos diferentes estudos [24, 28, 34] e a idade também com resultados variáveis [16, 19, 24, 34].

Tabela 2 – Variáveis incluídas no modelo final dos estudos considerados

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	MODELO UTILIZADO (MEDIDA DE RISCO)	VARIÁVEL(EIS) INCLUÍDAS NO MODELO FINAL	VARIAÇÃO DE RISCO (DE ACORDO COM O MODELO UTILIZADO)
Mocroft <i>et al.</i> , RU, 2001 [16]	Modelo de Cox com covariáveis dependentes do tempo (HR)	Idade	0,73 (0,60-0,88) (Por 10 anos)
		Experiência tratamento	0,65 (0,48-0,88) (Naïve vs. não)
		Número de fármacos no esquema	2,21 (1,50-3,26) (≥ 4 fármacos vs. <4 fármacos)
		Participação em ensaio clínico	0,64 (0,43-0,95)
		Esquema com RTV	1,41 (1,02-1,94)
		Esquema com NFV	0,57 (0,34-0,95)
		Carga vírica *	1,51 (1,34-1,70)
Li <i>et al.</i> , EUA, 2005 [19]	Regressão logística com controlo da dependência (OR)**	Idade	1,22 (1,00 – 1,48) (Diminuição de 5 anos)
		Carga vírica	1,42 (1,17-1,72) (Aumento de log ₁₀)
		Depressão (escala)	2,03 (1,24-3,32)
		Esquema com ABC	1,82 (1,03-3,20)

Tabela 2 (cont.) – Variáveis incluídas no modelo final dos estudos considerados

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	MODELO UTILIZADO (MEDIDA DE RISCO)	VARIÁVEL(EIS) INCLUÍDAS NO MODELO FINAL	VARIAÇÃO DE RISCO (DE ACORDO COM O MODELO UTILIZADO)
Kiguba <i>et al.</i> , Uganda, 2007 [28]	Regressão logística (OR)	Duração HAART	3,13 (1,16-8,33) (>3 meses em HAART vs. ≤3 meses)
		Estado civil	1,64 (1,02-2,70) (Não casado vs. casado)
		Tipo de regime	1,51 (0,80-2,85) (INNTR vs. INTR)
		Infeções Oportunistas	1,37 (0,83-2,25)
		Contagem linfócitos CD4 ⁺	1,28 (0,77-2,12) (≥200 vs. <200)
de la Torre <i>et al.</i> , Espanha, 2008 [12]	Modelo de riscos proporcionais de Cox (HR) ***	Regime	INNTR – 1,58 (1,0-2,50) INTR – 2,84 (1,51-5,33) (vs. IP)
		VHC negativo	1,91 (1,04-3,52)
Golub <i>et al.</i> , EUA, 2008 [32]	Modelo de riscos proporcionais para tempos discretos (OR)	SIDA pós-HAART	4,46 (1,39 – 14,25)
		Experiência terapêutica pré-HAART	1,69 (1,16-2,47)
		Carga vírica	1,40 (1,23-1,60) (por aumento de log10)
		História de tratamento psiquiátrico	1,85 (1,21-2,82)
		Número de consultas nos últimos 6 meses	2,29 (1,55-3,39) (≥ 3 consultas últimos 6 meses vs. <3 consultas)

Tabela 2 (cont.) – Variáveis incluídas no modelo final dos estudos considerados

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	MODELO UTILIZADO (MEDIDA DE RISCO)	VARIÁVEL(EIS) INCLUÍDAS NO MODELO FINAL	VARIAÇÃO DE RISCO (DE ACORDO COM O MODELO UTILIZADO)
Vo <i>et al.</i> , Suíça, 2008 [34]	Regressão logística multinominal (OR)	Sexo	1,22 (0,94-1,57) (Mulheres vs. homens)
		Idade	1,04 (0,93-1,17) (Por 10 anos)
		Consumo de drogas injetáveis	Passado ou sob metadona - 1,40 (0,79-2,49) Atual - 0,92 (0,54-1,58) (vs. nunca)
		HBV (Infecção ativa)	1,00 (0,60-1,64)
		HCV	0,77 (0,48-1,24)
		Diagnóstico de SIDA	1,01 (0,75-1,24)
		Contagem de linfócitos CD4+	<200 - 1,09 (0,83-1,43) ≥350 - 1,50 (1,04-2,17) (vs. 200-350)
		Carga vírica	1,35 (1,07-1,71) (>5log10 vs. ≤5log10)
		Fármacos contidos no esqueleto	D4T/3TC - 1,57 (0,93-2,66) TDF/ FTC ou 3TC - 0,65 (0,43-0,97)
			ddI/outro INTR - 2,06 (1,29-3,31) Outros - 1,21 (0,67-2,21) (vs. ZDV/3TC)
			LPV/r - 1,09 (0,81 – 1,45) NFV - 0,81 (0,57-1,15) IDV/r - 2,28 (1,24-4,17) NVP - 0,94 (0,49-1,79) ABC - 0,78 (0,37-1,61) ATV/r - 0,52 (0,22 – 1,26)
		3º fármaco	Outros IP sem RTV - 0,36 (0,13-1,03) Outros IP com RTV - 0,99 (0,34-2,84) Outros - 1,97 (0,84-4,58) (vs. EFV)

Tabela 2 (cont.) – Variáveis incluídas no modelo final dos estudos considerados

ESTUDO, LOCAL, ANO DE PUBLICAÇÃO	MODELO UTILIZADO (MEDIDA DE RISCO)	VARIÁVEL(EIS) INCLUÍDAS NO MODELO FINAL	VARIAÇÃO DE RISCO (DE ACORDO COM O MODELO UTILIZADO)
Landier <i>et al.</i> , 2010, Mali [24]	Modelo de riscos proporcionais de Cox (HR)	TAR prévia	3,8 (1,6-8,8)
			≤50 - 4,2 (1,2-15,0)
		Contagem de células CD4 ⁺	51-100 - 3,4 (0,9-13,1)
			101-200 - 1,9 (0,5-7,1)
			(vs. ≥200)
		Idade	30-39 anos - 2,2 (0,9-5,6)
			≥40 anos - 3,4 (1,3-8,8)
			(vs. <30 anos)
		INNTR (VIH2)	80,8 (21,8-299,5)

* Variável dependente do tempo

** Também ajustado pelo centro de estudo, OR respectivo não apresentado

*** Fatores associados à durabilidade do regime terapêutico inicial

Legenda: 3TC – lamivudina; ABC – abacavir; ATV/r – atazanavir potenciado por ritonavir; D4T – estavudina; ddi – didanosina; EFV – efavirenz; EUA – Estados Unidos da América; FTC – emtricitabina; HAART – Terapêutica antiretroviral de alta eficácia; HR – Razão de riscos; IDV/r – indinavir potenciado por ritonavir; INNTR – Inibidor não nucleósido da transcriptase reversa; INTR - Inibidor nucleósido da transcriptase reversa; IP – inibidor protease; LPV/r – lopinavir potenciado por ritonavir; NFV – nelfinavir; NVP – nevirapina; OR – Razão das chances; RTV – ritonavir; RU – Reino Unido; SIDA – Síndrome da imunodeficiência adquirida; TAR – Terapêutica antiretroviral; TDF – tenofovir; VHB – Vírus da hepatite B; VHC – Vírus da hepatite C; VIH-2 – Vírus da imunodeficiência humana tipo 2; ZDV – zidovudina.

Como visto anteriormente poucos são os trabalhos que avaliaram os fatores associados às várias mudanças terapêuticas. A nível nacional apenas são conhecidos trabalhos apresentados em congresso e nenhum deles avalia os fatores associados à mudança terapêuticas. Assim, é necessário estudar e analisar estes aspetos.

Objetivo

O objetivo do presente trabalho é descrever os fatores associados à mudança terapêutica nos doentes seropositivos para o VIH-1, utilizando os modelos de sobrevivência mais adequados às características do estudo.

Metodologia

O projeto

O presente trabalho integra-se no estudo *Adesão à Terapêutica AntiRetroviral em doentes adultos seropositivos para o VIH* (ATAR-VIH), desenvolvido pelo Instituto de Medicina Preventiva da Faculdade de Medicina de Lisboa. Trata-se de um estudo observacional de coorte retrospectiva (e uma fase prospetiva posterior) com o objetivo de determinar as variáveis associadas com a adesão à terapêutica antiretroviral, numa amostra de indivíduos seropositivos para o VIH-1 seguidos em ambulatório num hospital de nível três (Hospital de Santa Maria - HSM, Lisboa) [3].

Partindo dos registos dos indivíduos que tiveram dispensas de TAR, entre janeiro de 2005 e dezembro de 2008, foi selecionada uma amostra aleatória simples de 320 indivíduos. Seguiu-se a análise dos respetivos processos clínicos, de abril a setembro de 2010.

Os indivíduos foram considerados elegíveis se:

1. Tivessem pelo menos 18 anos no início da terapêutica;
2. Tivessem iniciado terapêutica no Hospital de Santa Maria;
3. Tivessem pelo menos duas consultas no período entre 1 de janeiro de 2005 e 31 de dezembro de 2009 no Serviço de Doenças Infecciosas (SDI) do referido hospital;
4. O início de TAR não tivesse sido motivado por medidas profiláticas;
5. Não tivessem participado em ensaios clínicos.

Foi consultado o processo clínico de cada indivíduo elegível, com recolha de informação desde o início do seguimento no SDI até 31 de dezembro de 2009 ou censura. Os critérios de censura foram:

- Falecimento;
- Início de dependência de terceiros para a toma da medicação;
- Data de detenção em prisão;
- Entrada em instituição social;
- Paragem na terapêutica, uma vez que alguns estudos realizados previamente apontam para diferenças nos fatores associados à mudança terapêutica e paragem [16, 19, 28, 34].

Partindo de uma base de dados, anonimizada, construída em Microsoft Office AccessTM, os dados foram exportados para o programa R (versão 2.15.0 - [www.r-](http://www.r-project.org)

project.org) para posterior análise estatística [37]. A análise começou por ser descritiva, utilizando-se frequências relativas e absolutas para as variáveis nominais e medidas de tendência central e de dispersão para as numéricas [11].

O projeto ATAR-VIH foi autorizado pela Comissão de Ética do Centro Hospitalar Lisboa Norte e pela Comissão Nacional de Proteção de Dados.

Variável resposta e evento de interesse

Na recolha dos dados dos processos clínicos podem ser considerados vários momentos:

- Início do seguimento no SDI – correspondente ao momento da primeira consulta no serviço, ainda sem terapêutica;
- Início de TAR – correspondente à data da consulta onde foi prescrita terapêutica;
- Início de HAART – correspondente à data de início de terapêutica com três ou mais fármacos. Corresponde à data de início de TAR caso a primeira terapêutica do indivíduo seja do tipo HAART.

Para a análise do evento de interesse – a mudança terapêutica – foram consideradas as mudanças de regimes HAART (definidos como três ou mais fármacos) [2]. Desta forma, os doentes foram considerados em risco de mudança apenas após o início de HAART. Todas as mudanças foram consideradas como representando o mesmo tipo de acontecimento.

Sendo utilizadas técnicas de análise de sobrevivência, a variável resposta corresponde ao tempo até à mudança terapêutica. Uma vez que se pretende modelar um evento que se repete no tempo as extensões do modelo de Cox para eventos repetidos no tempo (modelos de eventos múltiplos) são as mais adequadas.

Para a compreensão dos modelos de eventos múltiplos importa fazer uma revisão dos pressupostos teóricos subjacentes. Apresenta-se uma breve revisão do modelo de regressão de Cox, necessária à compreensão dos modelos de eventos múltiplos aqui analisados, seguindo-se então a apresentação dos referidos modelos, aplicados aos dados em questão.

Regressão de Cox

O modelo de regressão de Cox é o mais utilizado na modelação de variáveis com dados censurados [38].

Seja $Z_i(t)$ o vetor de p covariáveis para o i -ésimo indivíduo, que podem ser fixas ao longo do tempo ou variar ao longo do tempo, o modelo de Cox estima o risco para o indivíduo i como

$$\lambda_i(t) = \lambda_0(t)e^{Z_i'(t)\beta} \quad (1)$$

sendo $\lambda_0(t)$, a função de risco basal não especificada e β o vector de dimensão $p \times 1$ de parâmetros de regressão desconhecidos [38].

O modelo de regressão de Cox é por vezes chamado de modelo de riscos proporcionais vistos que a razão de riscos para dois indivíduos com vetores de covariáveis fixos Z_i e Z_f é constante ao longo do tempo, como mostra a expressão abaixo [38]:

$$\frac{\lambda_i(t)}{\lambda_f(t)} = \frac{\lambda_0(t)e^{Z_i'\beta}}{\lambda_0(t)e^{Z_f'\beta}} = \frac{e^{Z_i'\beta}}{e^{Z_f'\beta}} = e^{(Z_i' - Z_f')\beta}$$

A estimação dos coeficientes de regressão (β) é feita com base na verosimilhança parcial proposta por Cox:

$$L(\beta) = \prod_{k=1}^d \frac{\lambda(t_{(k)})}{\sum_{j \in R(t_{(k)})} \lambda(t_{(j)})}$$

em que d é o número de eventos observado. Sejam os tempos dos eventos ordenados representados por $t_{(1)} < t_{(2)} \dots < t_{(d)}$, em que $t_{(k)}$ é o k -ésimo tempo dos eventos ordenados. O grupo em risco respetivo, representado por $R(t_{(k)})$ é o conjunto de indivíduos em risco no tempo $t_{(k)}$.

Embora a verosimilhança parcial não seja, em geral, uma verdadeira verosimilhança pode ser tratada como tal para objetivos de inferência [38]. A estimativa dos coeficientes é então feita através do método da máxima verosimilhança parcial: é derivado o logaritmo da verosimilhança parcial ($U(\beta)$), que é depois igualado a zero. A equação obtida é depois resolvida iterativamente com recurso ao algoritmo de Newton-Raphson, implementado na biblioteca *survival* do programa R [37, 39]:

$$\hat{\beta}^{(n+1)} = \hat{\beta}^{(n)} + I^{-1}(\hat{\beta}^{(n)})U(\hat{\beta}^{(n)})$$

$\hat{\beta}^{(0)}$ é habitualmente 0 e $I^{-1}(\hat{\beta})$ corresponde à inversa da matriz de informação, que é usada como a variância de $\hat{\beta}$ [38].

Modelos de eventos múltiplos

Os modelos de eventos múltiplos surgiram como extensões do modelo de regressão de Cox, permitindo a modelação de dados cujo evento de interesse se repete para o mesmo indivíduo ou em que o mesmo indivíduo apresenta vários eventos de interesse [38, 40].

Vários foram os modelos propostos para a modelação deste tipo de eventos. No caso de eventos ordenados destacam-se três: o modelo de Andersen e Gill (AG), o modelo Prentice, Williams e Peterson (PWP) e o modelo de Wei, Lin e Weissfeld (WLW) [38, 40, 41]. Nestes modelos, os coeficientes são estimados ignorando-se a possível correlação existente entre os eventos (dado que um mesmo indivíduo pode contribuir com mais que uma observação e por isso estas não são independentes). A correção para a dependência faz-se através da estimação de uma variância corrigida, semelhante ao estimador sanduíche utilizado nas equações de estimação generalizadas – estimador ABA em que A corresponde à variância usual (a inversa da matriz de informação I) e B é um termo de correção [38, 42]. Embora tenham uma base comum, estes modelos apresentam diferenças subtis, conduzindo a diferentes resultados [40].

Modelo Andersen-Gill:

O modelo proposto por Andersen e Gill é considerado o mais simples dos modelos aqui discutidos. Contudo, é também o que apresenta mais pressupostos [38]. Embora surja como uma extensão do modelo de Cox, aproxima-se da regressão de Poisson, ao assumir que o evento a modelar segue um processo de Poisson dependente do tempo [38]. Assim, neste modelo admite-se que o risco de sofrer o evento se mantém independentemente dos eventos prévios, como se pode verificar pela Equação (2), que traduz a equação de risco para o indivíduo i .

$$\lambda_i(t) = Y_i(t)\lambda_0(t)e^{Z_i'(t)\beta} \quad (2)$$

Esta equação assemelha-se à função de risco para o modelo de Cox [Equação (1)]. Da análise e comparação das equações apresentadas [Equações (1) e (2)] podemos verificar que a diferença entre a função de risco do modelo de Cox e do modelo AG reside na definição de $Y_i(t)$ (a variável indicatriz de risco, ou seja, que toma o valor 1 se o indivíduo i se encontrar em observação e em risco no momento t e o valor 0 caso contrário). No modelo de regressão de Cox quando um evento ocorre o indivíduo já não

está sob risco ($Y_i(t) = 0$ para $t > t_{evento}$). Já no modelo AG $Y_i(t)$ pode ser 1 para $t > t_{evento}$, enquanto o indivíduo se encontrar em risco de sofrer novo evento.

De salientar que, embora neste modelo se admita classicamente que o risco de sofrer o evento se mantém para todos os eventos, este pressuposto pode ser relaxado pela introdução no modelo de covariáveis dependentes do tempo, como o número de eventos prévios, que poderão captar a estrutura de dependência entre os tempos de recorrência [41].

A verificação dos pressupostos do modelo AG deve ser feita pela comparação das estimativas das variâncias robusta e naïve, que deverão ser sobreponíveis [38].

Para proceder ao ajustamento deste tipo de modelos a construção da base de dados a analisar tem um papel fulcral. A base de dados para cada modelo irá variar no número de linhas total, na formulação do intervalo de risco e na existência de coluna de ordenação do evento (coluna de estrato). No caso do modelo AG a formulação do intervalo de risco segue uma estrutura de processos de contagem: (entrada no estudo, primeiro evento], (primeiro evento, segundo evento], ..., (j -ésimo evento, último registo] [38, 40]. Esta última linha existirá apenas se o indivíduo esteve em seguimento após o último evento. Caso o último evento coincida com o último registo não será incluída. Assim, o número de linhas do indivíduo i será dado por $M_i + D_i$, sendo M_i a função correspondente ao número de eventos de interesse para o indivíduo i e

$$D_i = \begin{cases} 1, & \text{se o indivíduo } i \text{ permanece em observação após o último evento} \\ 0, & \text{caso contrário} \end{cases}$$

O número total de linhas da base de dados é naturalmente o somatório do número de linhas de cada indivíduo - $\sum_{i=0}^n (M_i + D_i)$.

Um exemplo de aplicação deste tipo de modelo seria o estudo dos fatores associados a infeções respiratórias recorrentes, assumindo que uma nova infeção respiratória não altera o risco de uma infeção subsequente [43].

Modelo Wei-Lin-Weissfeld (ou modelo marginal)

Embora descrito em conjunto com os modelos que tratam de modelos de eventos múltiplos para eventos ordenados o modelo proposto por Wei, Lin e Weissfeld trata os

dados como se estivéssemos perante uma situação de riscos competitivos não ordenados [38, 44].

A Equação (3) traduz a função de risco que, pela formulação do modelo, é definida não só para cada indivíduo mas também para cada evento, visto que, ao contrário do que acontece no modelo AG, neste admite-se que o risco pode variar para cada evento.

$$\lambda_{ij}(t) = Y_{ij}(t)\lambda_{0j}(t)e^{Z_i'(t)\beta_j} \quad (3)$$

Tal como explicado anteriormente existem diferenças na construção da base de dados, relativamente ao modelo AG:

- O intervalo de tempo para a p -ésima linha de cada indivíduo é apresentado sob a forma de tempo total e não de processos de contagem: (entrada no estudo, primeiro evento], (entrada no estudo, segundo evento], ..., (entrada no estudo, último registo].
- É necessária a criação de uma coluna de estrato que corresponderá ao número do evento de cada indivíduo.
- O número de linhas para cada indivíduo é igual para todos e corresponde ao número máximo de eventos sofridos. Sendo J o número máximo de eventos a dimensão da base de dados será nJ .

Este tipo de modelo poderá ser aplicado em situações em que mais que um desfecho é possível, por exemplo, com dados de efeitos adversos de uma medicação, uma vez que mais que um efeito adverso poderá ser observado [45].

Modelo Prentice-Williams-Peterson (ou modelo condicional)

O modelo proposto por Prentice, Williams e Peterson em 1981 é também conhecido como modelo condicional, uma vez que assume que um indivíduo não está em risco de sofrer o j -ésimo evento até ter sofrido o evento $j - 1$ [46].

A função de risco é semelhante à apresentada para o modelo WLW diferindo na definição de $Y_{ij}(t)$: este será zero até ocorrer o evento $j - 1$, momento a partir do qual passa a ser um (definição que decorre dos pressupostos do modelo, como explicado acima) [38, 44].

Para ajustar um modelo deste tipo, a base de dados tem semelhanças com as anteriormente apresentadas. Por um lado aproxima-se da base descrita para o modelo AG, quer na formulação do intervalo em risco (através dos processos de contagem), quer no número total de linhas. Por outro, tal como no modelo WLW será necessária a criação da variável estrato, que corresponde ao número de ordem da ocorrência do evento para cada indivíduo [38].

Para além da formulação do intervalo de risco através de processos de contagem anteriormente descrita, o modelo PWP também pode ser estimado considerando o intervalo de risco com formulação *gap time*. Nesta formulação após cada evento o tempo recomeça do zero. A diferença em relação às formulações anteriores reside na definição do número de indivíduos em risco num determinado momento.

Na Figura 2 encontram-se resumidas de forma esquemática as possíveis formulações do intervalo de risco neste tipo de modelos. São apresentados três indivíduos: o indivíduo A que sofre dois eventos (tempo 5 e 10, respetivamente) e é censurado no tempo 16; o indivíduo B que sofre três eventos (nos tempos 7, 12 e 17) e o indivíduo C que não sofre qualquer evento, sendo censurado no tempo 12.

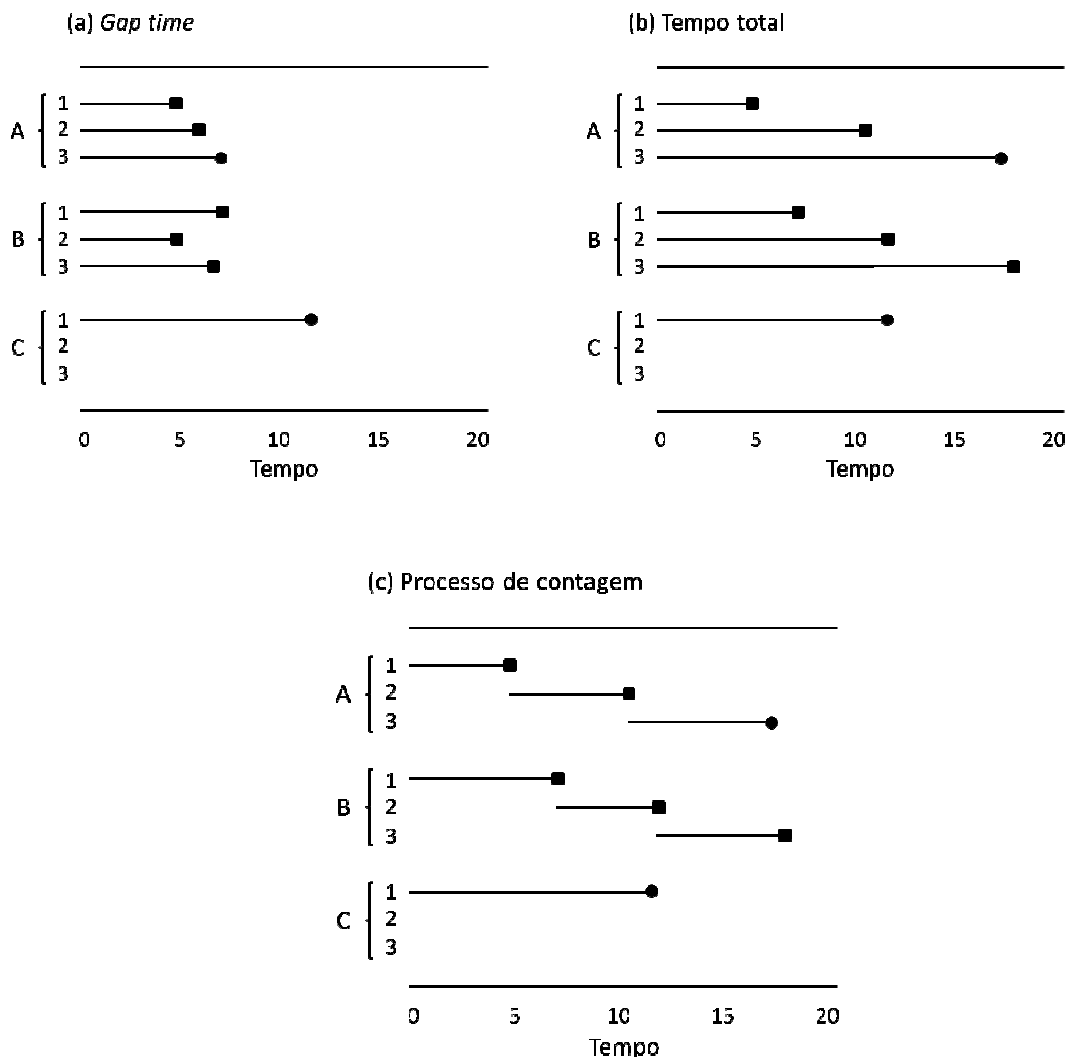


Figura 2 – Representação esquemática das possíveis formulações de intervalos de risco – (a) Gap time, (b) Tempo total, (c) Processos de contagem. O símbolo ■ representa a ocorrência de um evento e o ● o momento da censura (adaptado de Kelly *et al.* [40])

O modelo PWP poderá ser utilizado, para o estudo de fatores associados a eventos repetidos no tempo, em situações em que a ocorrência de um evento altere o risco da ocorrência do evento seguinte, como por exemplo em casos de diarreia infecciosa [43].

Resumo da utilização dos vários modelos

Kelly *et al.* propuseram a caracterização sistemática destes modelos de acordo com quatro componentes chave: intervalo em risco, risco basal, grupo em risco e ajustamento da correlação, sendo o intervalo em risco e o grupo em risco considerados como

essenciais para a construção do modelo [40]. Estas características permitem, segundo os autores, a escolha do modelo mais adequado aos dados em questão [40]. Essa sistematização é apresentada na Tabela 3. Considerando o tipo de risco, que pode ser comum a todos os eventos ou específico de cada um, como vimos anteriormente, os modelos podem ser divididos de acordo com o grupo em risco. Os modelos com risco basal comum têm um grupo em risco não restrito, enquanto que os modelos com função de risco específica de evento podem apresentar um grupo em risco semi-restrito ou restrito. Nos casos em que se considera o grupo em risco não restrito, o risco é comum a todos os eventos e todos os indivíduos contribuem para qualquer um dos eventos, independentemente do número de eventos que apresentam. Já no restrito só contribuem para o j -ésimo evento os indivíduos para quem o evento $j - 1$ já tenha ocorrido. Por fim, o semi-restrito aproxima-se do não restrito diferindo na possibilidade de apresentar riscos específicos de evento.

Tabela 3 – Definição do intervalo de risco e grupo em risco para cada modelo (adaptado de Kelly *et al.* [40])

CARACTERÍSTICAS DO MODELO		GRUPO EM RISCO/RISCO BASAL	
INTERVALO DE RISCO	Não restrito/comum	Semi-restrito/específico de cada evento	Restrito/específico de cada evento
Gap time	Possível (GT-UR)*	Impossível	PWP-GT
Tempo total	LWA**	WLW	Possível (TT-R)*
Processos de contagem	AG	Possível	PWP-PC

*Modelos hipotéticos, não explicitamente apresentados na literatura; **Modelo pouco desenvolvido na bibliografia consultada pelo que não é discutido no presente trabalho.

Legenda: AG – Modelo de Andersen-Gill; GT-UR – *Gap time* – não restrito; LWA – Modelo de Lee, Wei e Amato; PWP-PC - Modelo de Prentice, Williams e Peterson com formulação do intervalo de risco em processos de contagem; PWP-GT – Modelo de Prentice, Williams e Peterson com formulação do intervalo de risco do tipo *Gap time*; TT-R – Tempo total – restrito; WLW – Modelo de Wei, Lin e Weissfeld.

A Figura 3 esquematiza as potenciais formulações do risco basal salientando as diferenças entre os possíveis modelos e permitindo a comparação com o modelo clássico de Cox.

Já na Tabela 4 são apresentadas as verosimilhanças parciais de cada modelo bem como a função de risco. As verosimilhanças parciais aqui apresentadas são a base da estimação dos parâmetros de regressão.

Pelas características acima apresentadas o modelo que mais se aproxima dos dados do presente trabalho parece ser o PWP, uma vez que é lícito admitir que se trata de um grupo em risco restrito (só contribuindo para o j -ésimo evento os indivíduos que já tenham sofrido o evento $j - 1$) e é possível que o risco basal seja específico de evento. Desta forma foi este o modelo adotado neste trabalho.

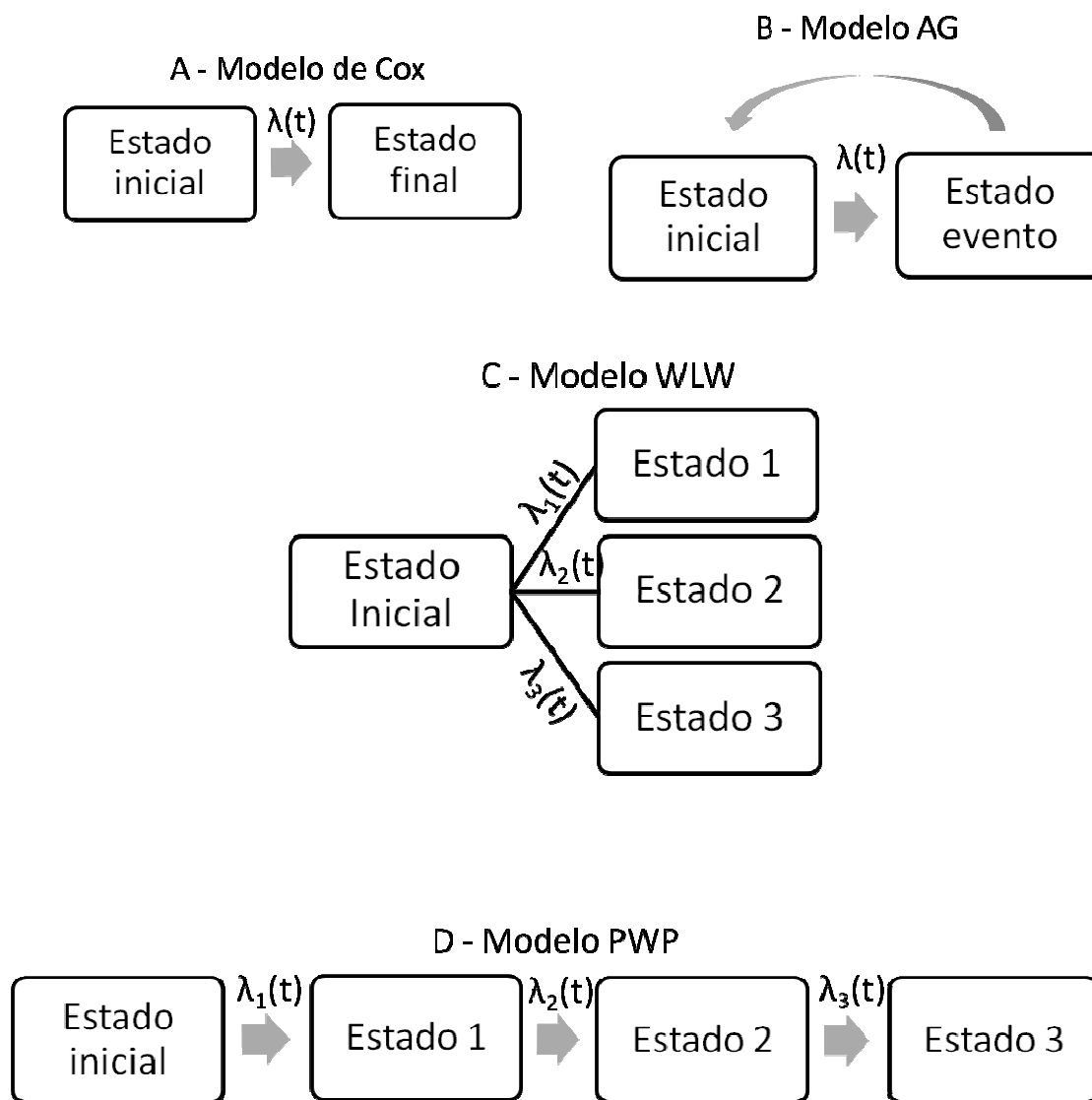


Figura 3 – Representação esquemática dos riscos basais no modelo de Cox (A) e nos vários modelos de eventos múltiplos (B – Modelo Andersen-Gill – AG; C – Modelo Wei, Lin e Weissfeld – WLW; D - Modelo de Prentice, Williams e Peterson – PWP). Adaptado de Carvalho *et al.* [43]

Tabela 4 – Equações das funções de risco e verosimilhanças parciais dos modelos apresentados (adaptado de Kelly *et al.* [40])

MODELO	VEROSIMILHANÇA PARCIAL	FUNÇÃO DE RISCO
AG	$L(\beta) = \prod_{i=1}^n \prod_{j=1}^J \left(\frac{e^{Z'_{ij}(X_{ij})\beta}}{\sum_{k=1}^n \sum_{l=1}^J Y_{kl}(X_{ij}) e^{Z'_{kl}(X_{ij})\beta}} \right)^{\delta_{ij}}$	$\lambda_{ij}(t; Z_{ij}) = \lambda_0(t) e^{Z'_{ij}(t)\beta}$ $Y_{ij}(t) = I(X_{i,j-1} < t \leq X_{ij})$
PWP-CP	$L(\beta) = \prod_{i=1}^n \prod_{j=1}^J \left(\frac{e^{Z'_{ij}(X_{ij})\beta}}{\sum_{k=1}^n Y_{kj}(X_{ij}) e^{Z'_{ij}(X_{ij})\beta}} \right)^{\delta_{ij}}$	$\lambda_{ij}(t; Z_{ij}) = \lambda_{0j}(t) e^{Z'_{ij}(t)\beta}$ $Y_{ij}(t) = I(X_{i,j-1} < t \leq X_{ij})$
PWP-GT	$L(\beta) = \prod_{i=1}^n \prod_{j=1}^J \left(\frac{e^{Z'_{ij}(X_{i,j-1} + G_{ij})\beta}}{\sum_{k=1}^n Y_{kj}(X_{ij}) e^{Z'_{ij}(X_{i,j-1} + G_{ij})\beta}} \right)^{\delta_{ij}}$	$\lambda_{ij}(t; Z_{ij}) = \lambda_{0j}(t - t_{j-1}) e^{Z'_{ij}(t)\beta}$ $Y_{ij}(t) = I(G_{ij} > t)$
WLW	Igual a PWP-CP	$\lambda_{ij}(t; Z_{ij}) = \lambda_{0j}(t) e^{Z'_{ij}(t)\beta}$ $Y_{ij}(t) = I(X_{ij} \geq t)$

Legenda: AG – Modelo de Andersen-Gill; PWP-CP – Modelo de Prentice, Williams e Peterson com formulação do intervalo de risco por processo de contagem; PWP-GT – Modelo de Prentice, Williams e Peterson com formulação do intervalo de risco por *gap time*; WLW – Modelo de Wei, Lin e Weissfeld. Notações: X_{ij} - tempo de observação do i -ésimo indivíduo, para o j -ésimo evento, δ_{ij} – variável indicatriz de censura (0 se censura, 1 caso contrário), $G_{ij} = X_{ij} - X_{i,j-1}$

Criação da base de dados para análise

Para o ajustamento deste tipo de modelos a construção da base de dados reveste-se de particular importância [38, 43]. Desta forma foi necessário proceder à transformação da base de dados existente para um formato compatível com a construção dos modelos pretendidos.

A base de dados inicial, construída em Microsoft Office AccessTM, continha informação dos vários indivíduos. Para cada um existiam várias consultas com uma variável identificativa da existência ou não de mudança, do tipo de mudança – mudança de pelo menos um dos fármacos, paragem de todo o regime ou reintrodução após uma paragem – e respetivo motivo. Abaixo estão representadas as linhas da base de dados de consultas (designada por *Consultas*), dos primeiros três participantes, e algumas das colunas de interesse.

No excerto apresentado estão assinalados alguns dos momentos chave para cada participante. A vermelho encontra-se o início do período em risco (início de terapêutica HAART), a azul encontram-se as mudanças de terapêutica consideradas e a verde o momento da censura.

```
> Consultas[1:58,c(1,2,30,31,39,40)]
```

id	DataConsulta	Mudanca	MudancaQual	TAR	TipoTAR
3	1993-10-21	0	NA	0	NA
3	1998-08-18	0	NA	1	HAART
3	2005-04-04	0	NA	1	HAART
3	2005-04-11	0	NA	0	NA
3	2005-09-23	1	Farmaco	1	HAART
3	2005-12-09	0	NA	1	HAART
3	2006-02-17	0	NA	1	HAART
3	2006-04-17	0	NA	1	HAART
3	2006-08-02	0	NA	0	NA
3	2007-04-20	0	NA	0	NA
3	2007-09-05	0	NA	1	HAART
3	2007-12-20	0	NA	1	HAART
3	2008-04-21	1	Posologia	1	HAART
3	2008-08-21	1	Posologia	1	HAART
3	2009-01-20	0	NA	1	HAART
3	2009-05-08	0	NA	1	HAART
3	2009-11-03	0	NA	0	NA

(cont.)

id	DataConsulta	Mudanca	MudancaQual	TAR	TipoTAR
6	1993-08-26	0	NA	0	NA
6	1996-02-12	0	NA	1	não HAART
6	1998-03-23	1	Farmaco	1	não HAART
6	2001-03-01	1	Farmaco	1	HAART
6	2002-02-10	1	Sem TAR	0	NA
6	2005-02-15	1	Reinicio/Farmaco	1	HAART
6	2005-05-11	1	Sem TAR	0	NA
6	2005-08-02	1	Reinicio/Farmaco	1	HAART
6	2005-09-28	0	NA	1	HAART
6	2006-03-23	1	Sem TAR	0	NA
6	2006-10-01	1	Reinicio de TAR	1	HAART
6	2006-12-01	1	Sem TAR	0	NA
6	2007-03-02	0	NA	0	NA
6	2007-07-04	0	NA	0	NA
6	2007-08-17	0	NA	0	NA
6	2007-11-08	0	NA	0	NA
6	2007-12-19	0	NA	0	NA
7	1992-02-06	0	NA	0	NA
7	1995-04-13	0	NA	1	não HAART
7	1995-07-13	1	Farmaco	1	não HAART
7	1999-01-11	1	Farmaco	1	HAART
7	2003-06-02	1	Farmaco	1	HAART
7	2005-04-06	0	NA	1	HAART
7	2005-06-24	1	Farmaco	1	HAART
7	2005-08-12	0	NA	1	HAART
7	2005-10-11	0	NA	1	HAART
7	2006-01-11	0	NA	1	HAART
7	2006-03-17	0	NA	1	HAART
7	2006-06-20	0	NA	1	HAART
7	2006-09-29	0	NA	1	HAART
7	2006-10-30	0	NA	1	HAART
7	2006-11-02	0	NA	1	HAART
7	2006-11-30	0	NA	1	HAART
7	2007-01-05	0	NA	0	NA
7	2007-03-06	0	NA	1	HAART
7	2007-06-08	0	NA	1	HAART
7	2007-11-21	0	NA	1	HAART
7	2008-04-17	0	NA	1	HAART
7	2008-09-17	1	Farmaco	1	HAART
7	2009-02-16	0	NA	1	HAART
7	2009-07-30	0	NA	1	HAART

A partir da base de dados apresentada foram implementados uma série de algoritmos no programa utilizado (cujo código pode ser consultado em anexo), que culminaram em duas base de dados – a utilizada para o ajustamento do modelo de Cox (PrimMudanca)

e a utilizada para o ajustamento do modelo PWP (`PWP1`). Os excertos das referidas bases de dados, para os três primeiros participantes, são também apresentados.

```
> PrimMudanca[1:3,1:7]1,2,3
```

id	dataHAART	dataUltima	status	ini	fim	tempo
3	1998-08-17	2009-11-03	1	26.3	112.7	86.4
6	2001-03-01	2002-02-10	0	57.2	68.7	11.5
7	1999-01-11	2009-07-29	1	31.2	84.6	53.4

```
> PWP1[1:7,c(1:7,22)]1,2,3
```

id	dataHAART	dataUltima	status	ini	fim	tempo	Estrato
3	1998-08-17	2009-11-03	1	26.3	112.7	86.4	1
3	1998-08-17	2009-11-03	0	112.7	162.8	50.1	2
6	2001-03-01	2002-02-10	0	57.2	68.7	11.5	1
7	1999-01-11	2009-07-29	1	31.2	84.6	53.4	1
7	1999-01-11	2009-07-29	1	84.6	109.7	25.1	2
7	1999-01-11	2009-07-29	1	109.7	149.1	39.4	3
7	1999-01-11	2009-07-29	0	149.1	159.6	10.5	4

Analisando a base de dados `Consultas`, para o participante 3, verifica-se que o período em risco se iniciou a 18-08-1998, tendo ocorrido uma mudança em 23-09-2005 e foi censurado a 03-11-2009, momento do último registo. Embora, para este participante, estejam registadas mais duas mudanças estas não foram consideradas para a presente análise. Tratam-se de alterações de posologia (número de comprimidos/dose a tomar por dia ou horário da toma), sendo que apenas as alterações de pelo menos um fármaco foram consideradas (designadas na base de dados por `Farmaco`). Correspondentemente na base de dados `PrimMudanca` verifica-se que ocorreu a primeira mudança (`status=1`), enquanto que na base de dados `PWP1` se verifica a existência de duas linhas para o referido participante. A primeira corresponde à primeira mudança e é equivalente à linha da base de dados `PrimMudanca`. Já a segunda traduz o tempo em risco após a ocorrência da mudança, tendo sido o participante censurado no final deste momento.

¹ Os tempos estão representados em meses e foram arredondados às décimas por questões de apresentação.

² Podem verificar-se diferenças de um dia na apresentação das datas (relativamente à base de dados `Consultas`) que se devem à conversão do formato de data no programa utilizado e não foram consideradas como alterando significativamente a análise dos dados.

³ `ini` e `fim` correspondem, respetivamente ao início e fim do intervalo em risco. `tempo` é a diferença aritmética dos dois.

O participante 6 apresenta apenas a consulta de início de HAART e de censura. Tal como explicado anteriormente, a paragem de terapêutica constituiu um dos critérios de censura no presente trabalho. Ainda que existam registos posteriores ao momento de censura, no participante em causa, a censura foi determinada pela paragem da terapêutica. Assim, quer na base de dados *PrimMudanca*, quer na *PWP1*, apenas se verifica uma linha, com *status=0* uma vez que o participante foi censurado no final do período em risco.

Para o participante 7 os registos assemelham-se aos do participante 3. Apenas se destacam diferenças no número de mudanças terapêuticas (mais duas que no caso do participante 3). Desta forma, a base de dados *PWP1* tem mais duas linhas que a do participante 3, correspondentes às mudanças observadas.

Covariáveis

Para a construção dos modelos foram consideradas as covariáveis disponíveis e passíveis de serem incluídas no modelo sem afetarem a qualidade dos dados (com baixo número de dados omissos). Essas variáveis foram:

- Sexo – variável categórica podendo ser considerados os níveis “Masculino” e “Feminino” [34].
- Idade – variável numérica, correspondendo à idade do indivíduo no início da terapêutica [16, 19, 24, 34].
- Experiência pré-HAART – variável categórica em que os doentes que tiveram regimes com menos de três fármacos tomam o nível “Sim” e os restantes o nível “Não” [32].
- Contagem de células $CD4^+$ - variável numérica correspondente ao valor de linfócitos T $CD4^+$ em circulação, cujo registo no processo fosse o mais próximo possível e até 3 meses antes do início da terapêutica HAART [24, 28, 34].
- Carga Vírica (CV) - variável numérica correspondente ao logaritmo de base 10 do valor de carga vírica registado no processo clínico, o mais próximo possível e até 3 meses antes do início da terapêutica HAART. Considerando que alguns indivíduos se encontravam sob terapêutica antes de iniciar HAART e que abaixo de 50 cópias/ml não é possível ter um valor sendo a CV designada de não detetável (ND) o valor constante foi ajustado ao zero. Assim,

$$\log_{10} CV_{considerado} = \log_{10}(CV_{registado} - 50)$$

De referir que uma vez que $\log_{10} 0 = -\infty$, para os valores de CV ND (e que depois de ajustados seriam 0) foi considerado o $\log_{10} 10^{-320} = -320$, próximo do limite permitido do programa R [16, 19, 32, 34].

- Regime HAART inicial – variável categórica que traduz os esquemas HAART iniciais [12, 28, 34]. Foram consideradas as combinações:
 - 2INTR+INNTR – no caso dos esquemas contendo dois fármacos da classe INTR e um da classe INNTR;
 - 2INTR+IP – no caso dos esquemas contendo dois fármacos da classe INTR e um da classe IP (potenciado ou não por ritonavir);
 - Outros – no caso dos esquemas contendo três fármacos da classe INTR ou quatro fármacos de qualquer classe (exceto IP/r).
- Consumo de drogas injetáveis – variável categórica em que os indivíduos com registo (prévio ou na consulta de início de HAART) de consumo de drogas injetáveis ou que tinham como forma de transmissão da infeção a utilização de drogas injetáveis tomavam o nível “Sim”, tendo o nível “Não” caso contrário [34].
- Ansiedade/Depressão – variável categórica traduzindo a existência de ansiedade/depressão. Aos participantes que tiveram registo da existência de Ansiedade/Depressão ou de prescrição de fármacos utilizados para o tratamento dessas patologias foi-lhes atribuído o nível “Sim” e aos restantes o nível “Não” [19, 32].
- Infeções oportunistas – variável categórica tomando o nível “Sim” para os participantes que tivessem tido até à data de início de HAART (inclusive) uma infeção oportunista e tomando o nível “Não” para os restantes [28].
- Neoplasias – variável categórica com as mesmas características que a anterior mas adaptada ao contexto neoplásico (neoplasias potencialmente relacionadas com a infeção VIH/SIDA).
- Período – Para além das variáveis apresentadas foi ainda criada uma variável para tentar controlar o facto de o padrão de prescrição ter evoluído ao longo do tempo, bem com as recomendações, o que poderia afetar os resultados do modelo final. Esta variável corresponde ao ano em que a terapêutica HAART foi iniciada e os indivíduos foram considerados como fazendo parte de três

períodos distintos – antes do ano 2000, entre 2000 e 2004 e de 2005 em diante. A escolha destes períodos relaciona-se com algumas alterações que poderão ter tido impacto, nomeadamente após 2000 por terem sido reduzidos os IP e a partir de 2005 por terem sido iniciada a utilização de fármacos em associação fixa (mais do que um princípio ativo num mesmo comprimido) [3, 32].

As referências apresentadas em cada variável correspondem a estudos em que os modelos finais incluíram as respetivas variáveis, tendo sido consideradas estatisticamente significativas ou não (exceto para a variável Período em que as referências sustentam a divisão efetuada).

Seleção de covariáveis

Foram construídos dois modelos:

1. Modelo de Cox para a primeira mudança,
2. Modelo PWP com variáveis no início do estudo.

A construção dos modelos assentou num processo *stepwise*. As variáveis foram incluídas caso o seu valor-p fosse inferior ao valor-p de inclusão ($P_E = 0,20$) e excluídas caso o valor-p fosse superior ao valor-p de exclusão ($P_S = 0,25$) [47]. Para o modelo de Cox o teste utilizado para a inclusão e exclusão de covariáveis foi o teste de razão de verossimilhanças [43]. Já no modelo PWP, e uma vez que o teste de razão de verossimilhanças assume independência das observações, foi utilizado o teste de Wald [39].

De referir ainda que no caso do modelo PWP foram consideradas apenas as mudanças até à quarta mudança, visto que após essa se verificou um número muito reduzido de mudanças o que dificultaria a estimação para os eventos dessa ordem [40].

Os modelos iniciais foram comparados com o modelo contendo a variável Período, para garantir que os resultados se encontravam controlados pelo período em que foi iniciada a terapêutica.

Construção do modelo – abordagem no software

Todos os modelos foram ajustados a partir da função *coxph()* da biblioteca *survival* [39]. Tratando-se de uma coorte aberta e uma vez que para além do tempo sob terapêutica também o período em que iniciavam e terminava a mesma (por censura ou mudança) poderiam ser importantes, o modelo de Cox foi ajustado com base na nomenclatura dos processos de contagem (ver Anexo):

```
coxph(Surv(ini,fim,status)~variáveis, ...)
```

Relativamente ao ajustamento do modelo de eventos múltiplos foi feito com recurso à mesma função do programa, com a inclusão de termos adicionais: por um lado o termo *strata* (*PWP\$Estrato*), que permite que o risco basal varie de acordo com a variável indicadora de estrutura ordenada dos eventos [43] e por outro, o termo *cluster* (*PWP\$idPARTICIPANTE*), que ao indicar que as linhas com o mesmo número de identificação pertencem à mesma pessoa vai levar ao cálculo da variância robusta que permite a correção da estrutura de dependência.

Para a execução do procedimento automático de seleção de covariáveis *stepwise* foi desenvolvido um algoritmo, cujo código se encontra em anexo, através do teste de razão de verosimilhanças e do teste de Wald, para o modelo de Cox e PWP, respetivamente.

Diagnóstico do modelo

Para o diagnóstico do modelo foram utilizadas medidas de ajustamento global e análise de resíduos.

Nas medidas de ajustamento global incluem-se o R^2 e a probabilidade de concordância, que são automaticamente fornecidos pelo programa utilizado. Se por um lado o R^2 pode ser interpretado como o poder explicativo das covariáveis no tempo de ocorrência do evento em estudo, por outro a probabilidade de concordância dá-nos noção do poder discriminativo do modelo [43].

Para a verificação do pressuposto de proporcionalidade das várias variáveis estudadas foram analisadas as curvas de sobrevivência (Kaplan-Meier) para cada uma das variáveis nominais, bem como realizado um teste de proporcionalidade global (através da função *cox.zph()*) [43]. Foram ainda realizados gráficos dos resíduos de Schoenfeld contra o tempo, para cada um dos coeficientes estimados pelo modelo final [43].

Os gráficos dos desvios residuais e martingala contra o índice de cada observação foram construídos e analisados com o intuito de identificar pontos aberrantes [43].

Finalmente analisaram-se os gráficos dos resíduos *score* contra as covariáveis incluídas no modelo final com o objetivo de se identificarem ponto influentes [43].

Resultados

Características gerais da amostra em estudo

De um total de 320 participantes, foram incluídos no estudo 194 (60,6%). Dos excluídos, 42 (33,3%) foram excluídos por terem iniciado terapêutica fora do HSM, 34 (27,0%) por terem participado em ensaios clínicos, 17 (13,5%) por terem menos de 2 consultas no SDI e 11 (8,7%) por terem iniciado terapêutica em contexto profilático.

As características gerais da amostra foram previamente descritas em detalhe [11]. Na Tabela 5 apresentam-se as características com maior relevância para o presente trabalho. Destaca-se o facto de a amostra ser formada maioritariamente por homens (63,9%) jovens (idade média $35,3 \pm 10,1$ anos) e caucasianos (76,8%).

Verifica-se que a transmissão heterossexual foi a via principal de aquisição da infeção (50,5%). De salientar que a categoria *Homossexual* (19,6%), inclui a categoria homo e bissexual (tal como definido pelo sistema de vigilância Português) enquanto a categoria *Outros* inclui situações em que apenas estava identificada a via sexual (sem menção a ser homo, hetero ou bissexual), casos de transmissão por produtos derivados do sangue, indivíduos que tinham mais do que uma possível via de transmissão ou os casos que o médico referiu explicitamente não ter sido identificada uma via de transmissão [48].

Tabela 5 – Características sociodemográficas, laboratoriais e de terapêutica dos 194 participantes da amostra estudada

n = 194 (salvo indicação contrária)	
<i>Características sociodemográficas</i>	
Idade média±DP – Início do seguimento médico	35,3±10,1 anos
Idade média±DP – Início de TAR	36,1±10,0 anos
Nº de mulheres	36,1%
Situação profissional no início do seguimento médico (dados omissos em 18,6%)	n = 158
Empregado	68,4%
Desempregado	20,3%
Outros	11,4%
Etnia (dados omissos em 22,2%)	n = 151
Caucasiana	78,8%
Negra	17,9%
Outras	3,3%
Transmissão (dados omissos em 5,2%)	n = 184
Heterossexual	49,5%
Utilizadores drogas injetáveis	22,3%
Homossexual	19,6%
Outros	8,7%

Tabela 5 (cont.) – Características sociodemográficas, laboratoriais e de terapêutica dos 194 participantes da amostra estudada

n = 194 (salvo indicação contrária)	
Características do indivíduo e da infeção	
Antecedentes de Infeções Oportunistas	37,1%
Antecedentes de Neoplasias	7,8%
Antecedentes de Ansiedade/Depressão	12,9%
Consumo de Drogas	25,8%
Características de diagnóstico e laboratoriais	
Ano de diagnóstico, mediana (IIQ)	2001 (1997 - 2003)
Meses de seguimento, mediana (IIQ)	23,1 (7,6 – 52,3)
Período	
Antes de 2000	29,4%
2000-2004	40,2%
2005 em diante	30,4%
Parâmetros laboratoriais	
Início de HAART	
Linfócitos T CD4+ - média±DP [casos omissos em 0,5%]*	241,3±203,9
Carga vírica considerada – média±DP [casos omissos em 4,6%]**	-2,2±47,4
Características da terapêutica	
Primeiro esquema HAART	
2 INTRs + INNTR	49,0%
2 INTRs + IP (potenciado por ritonavir ou não)	46,9%
3 INNTRs	3,6%
4 fármacos	0,5%
Início em regime não-HAART	22,2%

*células/ μ L, **cópias/mL

Legenda: DP, Desvio-padrão; IIQ - Intervalo interquartil; INNTR - Inibidor não-nucleósido da transcriptase reversa; INTR - Inibidor nucleósidos da transcriptase reversa; IP – Inibidor da protease; TAR - Terapêutica Antiretroviral.

Na figura 4 encontra-se o *box plot* relativo ao ano de diagnóstico que completa a informação apresentada na tabela 5. Verifica-se que em metade dos indivíduos estudados a infeção foi diagnosticada em 2001 ou antes, tendo por isso tempo de seguimento de vários anos (metade dos indivíduos como mais de 23 meses de seguimento). O mesmo fenómeno se verifica através da figura 5, onde se encontra representado o tempo de seguimento dos indivíduos. Neste caso considera-se tempo de seguimento o tempo em risco de sofrer o evento. Esta variável é assimétrica, dada a existência de tempos em risco bastante longos. De salientar, contudo, que apenas se verifica a existência de um *outlier*.

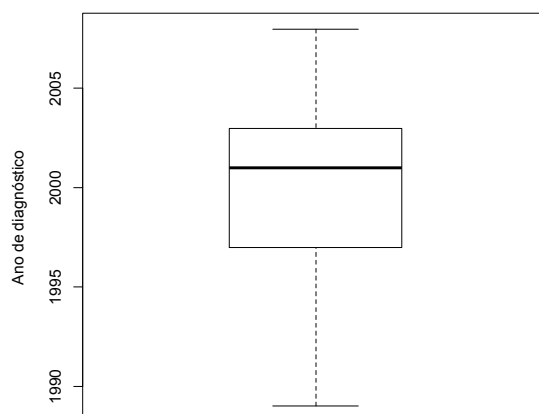


Figura 4 – Representação gráfica do ano de diagnóstico

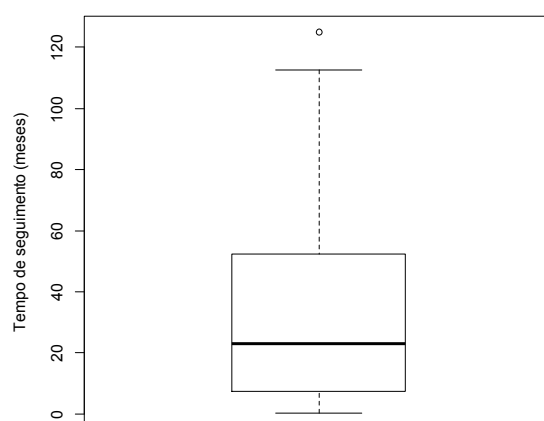
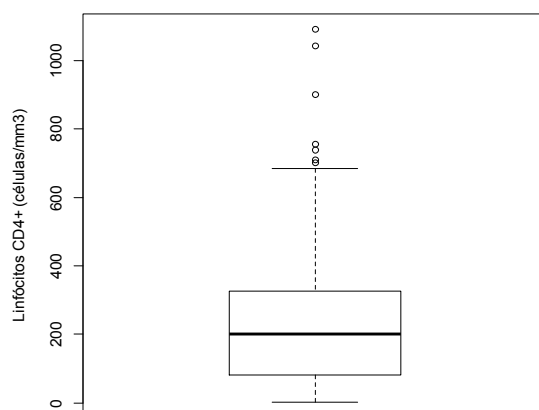


Figura 5 – Representação gráfica do tempo de seguimento, em meses

Na Figura 6 é ainda possível observar a tradução gráfica dos dados relativos aos valores dos linfócitos $CD4^+$. Verifica-se alguma assimetria e a existência de *outliers*. Já na Figura 7 encontra-se a representação gráfica dos valores de carga vírica considerados, à exceção dos valores correspondentes a carga vírica não detetável. Verifica-se também a existência de alguns *outliers*.



**Figura 6 – Representação gráfica dos valores de linfócitos $CD4^+$
no início da terapêutica HAART**

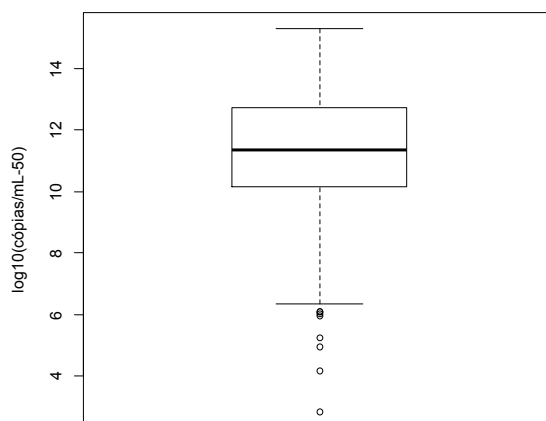


Figura 7 – Representação gráfica dos valores de carga vírica de acordo com a transformação utilizada. Os valores dos indivíduos que apresentavam carga vírica não detetável não se encontram representados (n=4) dado o valor de carga vírica considerado (-320) implicar diminuição de visibilidade no restante gráfico

Relativamente às mudanças terapêuticas, verificou-se que 100 indivíduos tiveram pelo menos uma mudança terapêutica, 57 pelo menos duas, 34 pelo menos três e 15 pelo menos quatro. O máximo de mudanças foram sete, sendo que não foram consideradas as mudanças após a quarta como anteriormente explicado (cf. Metodologia).

Considerando as variáveis de interesse e os dados omissos, foram incluídos um total de 185 indivíduos para o ajustamento dos modelos.

Modelo para o tempo até à 1ª mudança terapêutica

Análise Exploratória

Previamente ao ajuste do modelo foram desenhados as curvas de Kaplan-Meier considerando as variáveis nominais candidatas a incluir no modelo (Figura 8). Parece razoável assumir a proporcionalidade de riscos uma vez que mesmo nas variáveis Sexo, Regime, Consumo de drogas, antecedentes Psiquiátricos (Depressão/Ansiedade) e de infeções oportunistas se verifica uma sobreposição dos vários níveis.

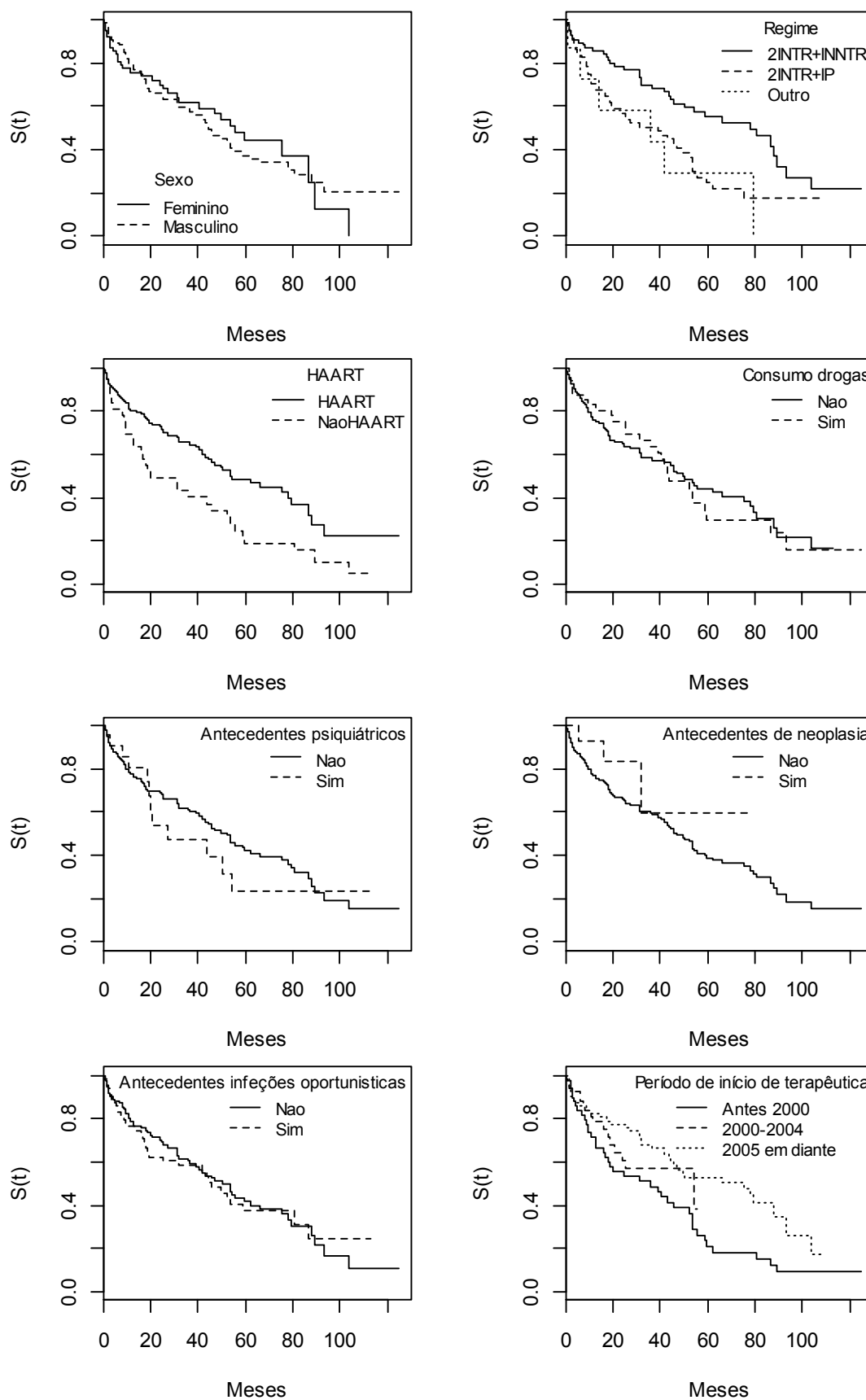


Figura 8 – Curvas de sobrevivência de Kaplan-Meier para as variáveis em estudo

Ajustamento do modelo

Na Tabela 6 são apresentados os valores-p resultantes do processo de seleção de covariáveis para o modelo de Cox. Verifica-se que a seleção do modelo implicou três passos, sendo que nenhuma das variáveis incluídas foi excluída nos passos posteriores à sua inclusão no modelo. As variáveis incluídas no modelo final foram o regime terapêutico inicial (Regime) e a experiência pré-HART (HAART).

Tabela 6 – Seleção de covariáveis pelo método *stepwise* (modelo de Cox)

PASSO	REGIME	HAART	IDADE	NEO	CD4	PSI	DROGAS	LOGCV	IO	SEXO
1	0,021	0,180	0,312	0,374	0,460	0,489	0,676	0,693	0,825	0,839
2	0,021*	0,133	0,765	0,287	0,700	0,297	0,548	0,771	0,935	0,974
3	0,017	0,133*	0,529	0,310	0,506	0,404	0,480	0,985	0,735	0,973

* Valores-p máximos de cada variáveis

Legenda: IO – Infecções Oportunistas, logCV – logaritmo da carga vírica, Neo – neoplasias, Psi – Ansiedade/Depressão

Tabela 7 – Razões de riscos e respetivos intervalos de confiança a 95% de acordo com o modelo de Cox escolhido

VARIÁVEIS	RAZÃO DE RISCOS (INTERVALO DE CONFIANÇA A 95%)
Período	
Antes de 2000	1
2000-2004	0,580 (0,296 – 1,133)
Depois de 2005	0,891 (0,358 – 2,219)
Regime	
2INTR + INNTR	1
2INTR + IP	1,777* (1,121 – 2,816)
Outro	2,587* (1,075 – 6,230)
Pré-HAART	
Não	1
Sim	1,520 (0,884 – 2,613)

* valor-p<0,05

Legenda: INNTR – inibidor não nucleósido da transcriptase reversa;
INTR – inibidores nucleósido da transcriptase reversa; IP – inibidor da protease.

As razões de risco e respetivos intervalos de confiança a 95% obtidos pelo ajustamento do modelo final são apresentadas na Tabela 7 (representada acima). Das variáveis incluídas apenas o regime terapêutico inicial está independentemente associado à

mudança. Quer os regimes constituídos por 2INTR+IP, quer os outros regimes, comparativamente aos regimes constituídos por 2INTR+INNTR aumentam o risco de mudança terapêutica, de forma estatisticamente significativa.

Qualidade de ajustamento

Das medidas da qualidade global de ajustamento, o R^2 obtido para o modelo final foi 7,5%, enquanto a probabilidade de concordância do modelo final foi de 60,1%.

Diagnóstico do modelo

Da análise dos gráficos contidos na Figura 9 podemos verificar que parece existir proporcionalidade ao longo do tempo. De salientar que, para a variável regime 2INTR+IP, os resíduos traduzem uma forma oscilatória mas com o coeficiente sempre contido no intervalo de confiança estimado. Embora nos restantes se verifique alguma variação esta pode ser atribuída à flutuação aleatória, não sendo valorizadas dado que não existe uma tendência clara na forma do gráfico.

Os aspetos descritos estão em concordância com os resultados do teste de correlação linear dos resíduos das covariáveis, no qual não se rejeita a hipótese de proporcionalidade dos riscos como se verifica pelo *output* abaixo (valor-p entre 0,755 para o Regime 2INTR+IP e 0,999 para o modelo global).

```
> cox.zph(modeloFinal)
```

	rho	chisq	p
Periodo+2005	-0.00373	0.00138	0.970
Periodo2000-2005	-0.00592	0.00367	0.952
Regime2INTR+IP	-0.03179	0.09710	0.755
RegimeOutro	0.01076	0.01109	0.916
HAARTNaohAART	0.02235	0.04272	0.836
GLOBAL	NA	0.20670	0.999

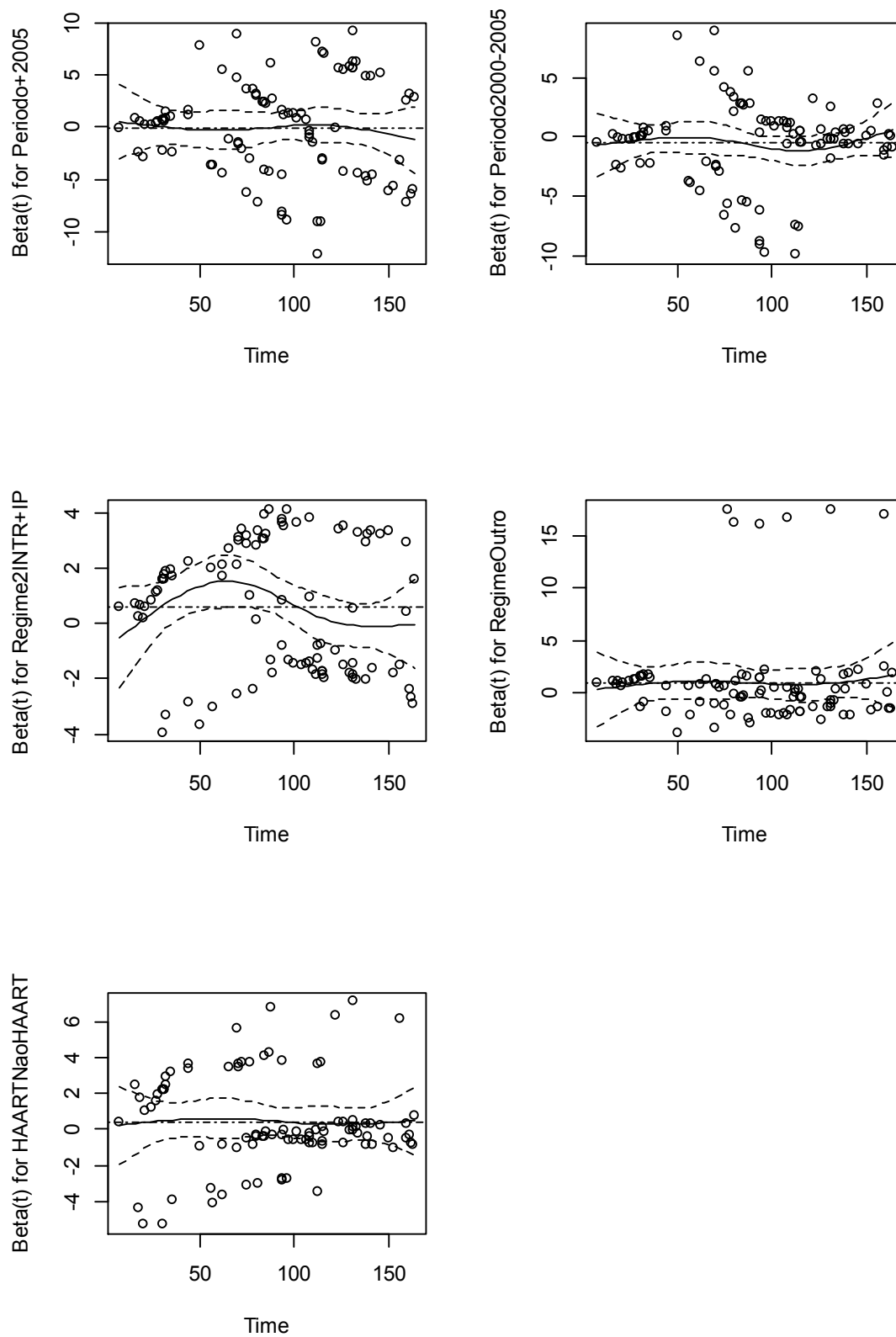


Figura 9 – Resíduos de Schoenfeld para o modelo final

Os gráficos dos resíduos de martingala e desvios residuais contra o índice, e dos desvios residuais contra os valores preditos tomam valores dentro do esperado, estando no intervalo $[-2, 2]$ (Figura 10, A a C). Relativamente ao gráfico quantil-quantil dos desvios residuais verifica-se que a distribuição dos resíduos se sobrepõe à esperada (Figura 10 – D).

A análise da Figura 11 permite analisar a existência de pontos influentes na construção do modelo [43]. Embora se encontrem marcados nos *box plots* alguns *outliers* verifica-se que a escala dos resíduos é muito reduzida e por isso não se considera a existência de pontos influentes.

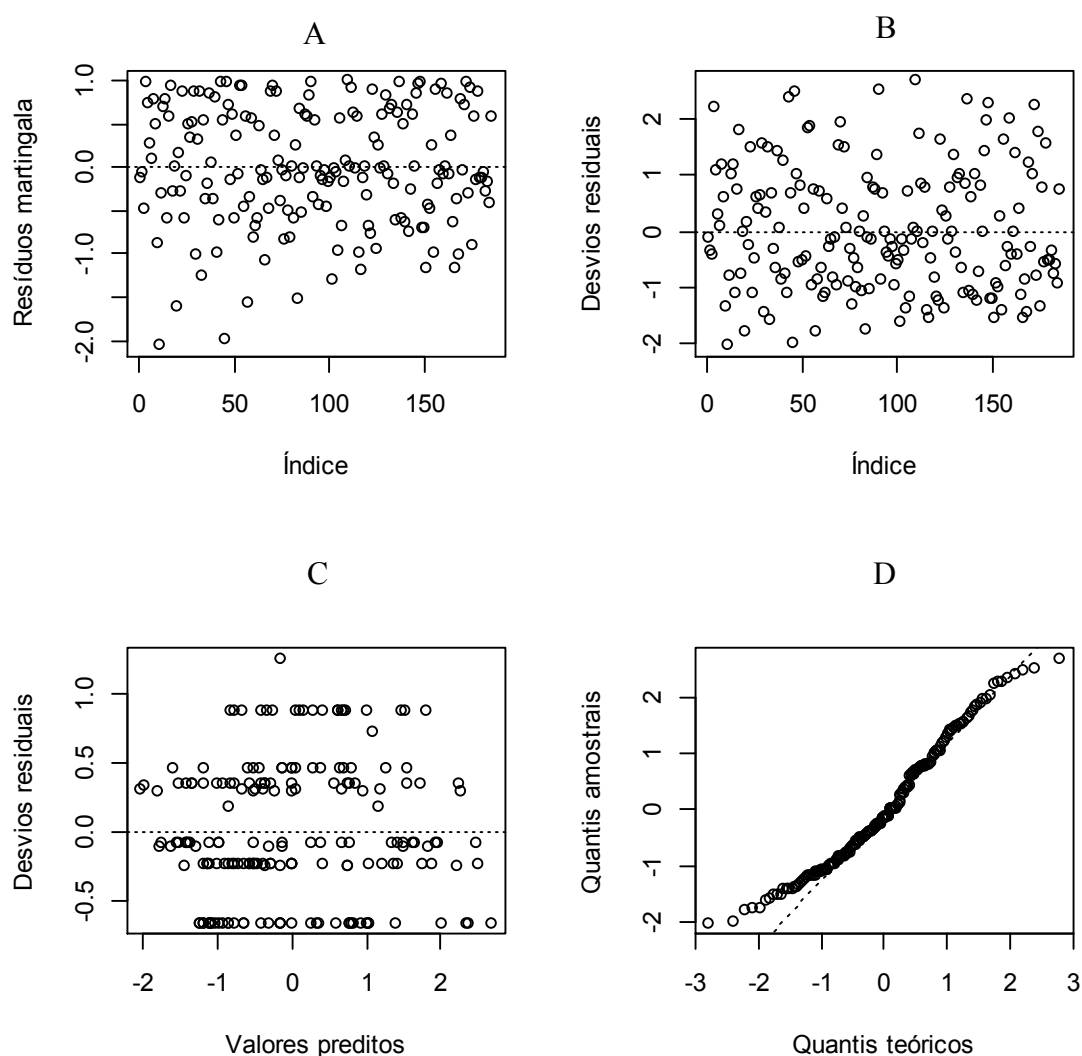


Figura 10 – Representação gráfica dos resíduos martingala (A) e desvios residuais (B) contra o índice, desvios residuais contra os valores preditos (C) e gráfico quantil-quantil dos desvios residuais (D)

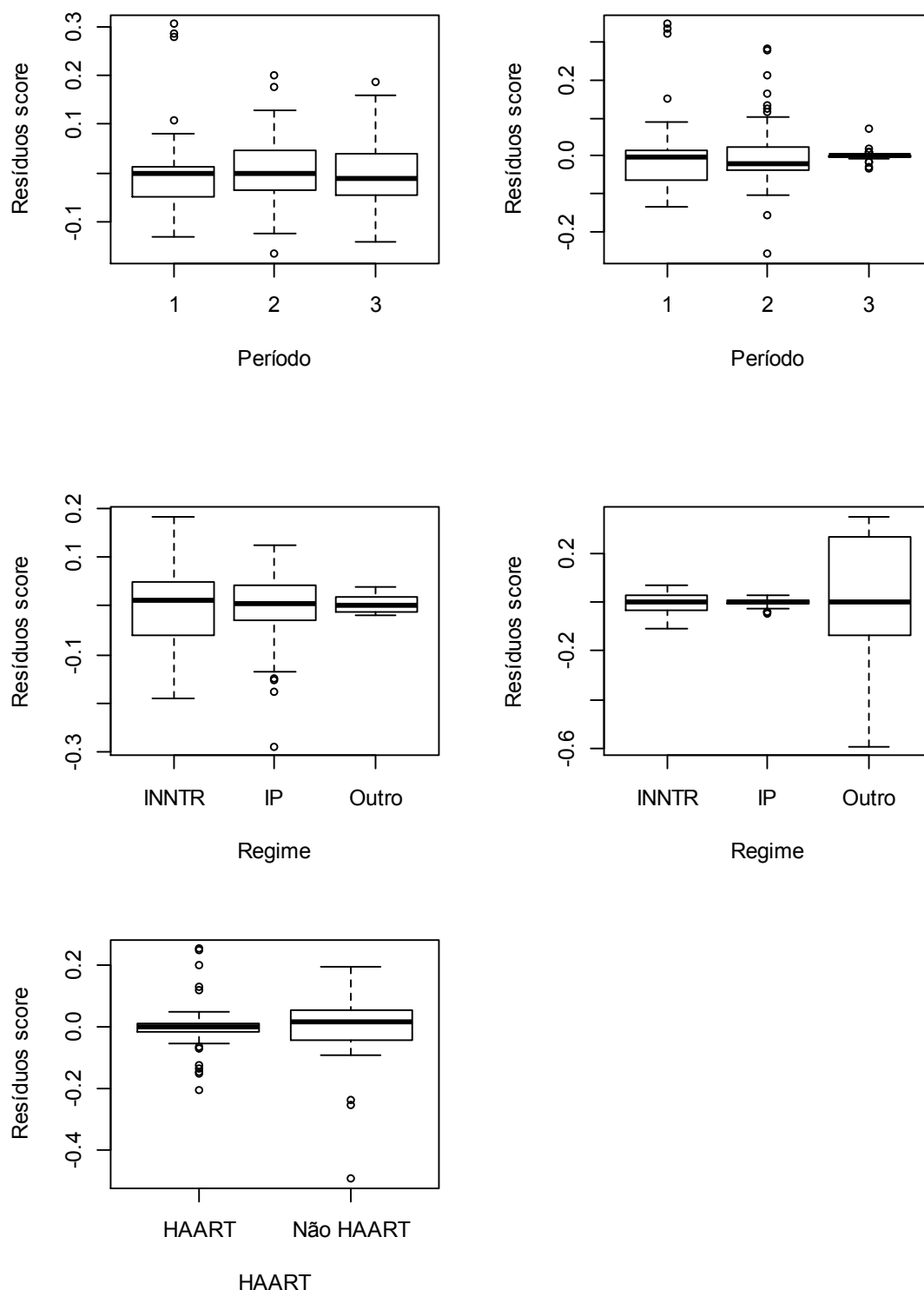


Figura 11 – Resíduos *score* contra as variáveis incluídas. Na variável Período 1, 2 e 3 refere-se ao período antes de 2000, entre 2000 e 2004 e após 2005, respetivamente. Na variável Regime, INNTR refere-se aos regimes contendo 2 INTR+INNTR e IP aos regimes com 2INTR+IP, respetivamente

Modelo para múltiplas mudanças terapêuticas

Para proceder ao ajustamento do modelo selecionado (PWP) para caracterização dos eventos utilizou-se, tal como anteriormente, o método *stepwise* com um valor-p de inclusão de 0,20 e de exclusão de 0,25. Na Tabela 8 são apresentados os valores-p resultantes do processo de seleção de covariáveis do mesmo. Verifica-se que a seleção do modelo decorreu em apenas dois passos levando à inclusão de uma variável - contagem de linfócitos CD4⁺.

Tabela 8 – Seleção de covariáveis candidatas pelo método *stepwise* (modelo PWP)

PASSOS	CD4	REGIME	HAART	IDADE	NEO	IO	LOGCV	DROGAS	SEXO	PSI
1	0,040	0,103	0,279	0,279	0,386	0,549	0,625	0,663	0,672	0,679
2	0,040	0,210	0,367	0,94	0,792	1,000	1,000	1,000	1,000	0,995

Legenda: IO – Infecções Oportunistas, logCV – logaritmo da carga vírica, Neo – neoplasias, Psi - Ansiedade/Depressão

As razões de risco e respetivos intervalos de confiança a 95% obtidos pelo ajustamento do modelo final são apresentadas na Tabela 9. Apenas a contagem de linfócitos CD4⁺ foi associada de forma independente às várias mudanças terapêuticas.

Tabela 9 – Razões de riscos e erro padrão das variáveis incluídas no modelo ajustado pelo método PWP

VARIÁVEIS	RAZÃO DE RISCOS (INTERVALO DE CONFIANÇA A 95%)	ERRO PADRÃO NAÏVE	ERRO PADRÃO ROBUSTO
Período			
2000-2004	0,716 (0,468-1,096)	0,205	0,217
Depois 2005	0,820 (1,400-1,531)	0,334	0,319
CD4	0,999* (0,998 – 0,9997)	0,0004	0,0004

*Valor-p<0,05

Na Figura 12 apresenta-se o risco acumulado de cada uma das mudanças estudadas, considerando o modelo final. Verifica-se que a segunda, terceira e quarta mudanças têm um perfil idêntico entre si. Relativamente à primeira mudança há uma desaceleração do ritmo de mudança a partir do mês 35.

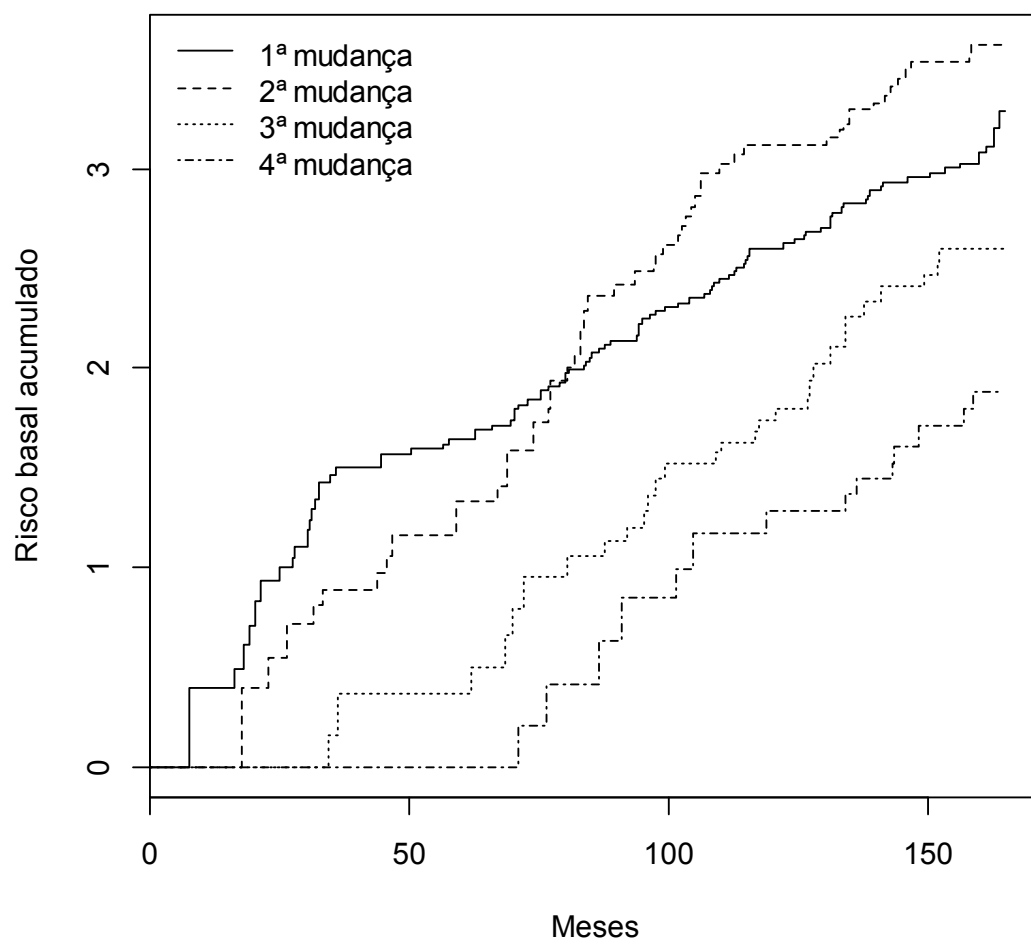


Figura 12 – Risco acumulado para a ocorrência de cada mudança

Discussão

Através do modelo de regressão de Cox e do modelo PWP foi possível estudar os fatores associados à primeira mudança terapêutica e às mudanças terapêuticas até à quarta. Abaixo segue-se a discussão dos resultados obtidos para cada um dos referidos modelos.

Modelo de Cox

No modelo de Cox final a única variável associada independentemente ao tempo até à primeira mudança terapêutica foi o regime terapêutico inicial. O modelo incluiu também as variáveis Período e experiência HAART prévia.

Analisando em maior pormenor cada uma das variáveis incluídas no modelo, verificamos que a variável Período não é estatisticamente significativa mas a sua inclusão no modelo permite que a inclusão das restantes variáveis não seja influenciada por prováveis diferenças no padrão de prescrição, que evoluiu ao longo dos anos [3, 32]. A idade foi, em alguns estudos anteriores, associada à mudança terapêutica. Por exemplo, nos estudos levados a cabo por Li *et al.* [19] e Landier *et al.* [24] o aumento da idade foi considerado um fator de risco, enquanto Mocroft *et al.* [16] apresenta a idade como um fator protetor para a mudança terapêutica. Por outro lado, Vo *et al.* [34] não encontraram uma associação estatisticamente significativa com esta variável, apesar da sua inclusão no modelo final. Neste trabalho, a variável idade não foi incluída no modelo final. A não inclusão poderá dever-se a diferenças sociodemográficas nas características das amostras.

O sexo dos participantes não foi incluído no modelo final. Em alguns estudos metodologicamente semelhantes ao presente trabalho (Tabela 2), somente um dos sexos foi estudado, o que dificulta comparações [19, 32]. Ainda assim, e considerando os resultados do presente e de estudos prévios, o sexo parece não estar associado à mudança terapêutica em geral.

Nesta análise o consumo de Drogas não foi considerado como estando associado à primeira mudança terapêutica. Noutros trabalhos verificou-se que este comportamento se encontrava associado à interrupção da terapêutica [34] ou à mudança por motivos específicos [33]. Assim o consumo de drogas não parece estar associado à mudança terapêutica quando são consideradas as mudanças no geral, apesar de ter sido associado à interrupção terapêutica ou à mudança por motivos específicos [33, 34].

Verificou-se um maior risco de mudança para os indivíduos que iniciaram terapêutica com regimes contendo IP (cerca de 77,8%) ou outros regimes, relativamente aos que

iniciaram terapêutica com INNTR. Nos caso dos IPs esta diferença tinha sido relatada anteriormente por de la Torre *et al.* [12] e Gratacòs *et al.* [49], podendo dever-se a um perfil de efeitos adversos menos favorável para os IPs. Contudo, os IPs mais recentes apresentam um perfil mais favorável sendo necessário avaliar se esta diferença se mantém atualmente ou não. Relativamente ao grupo dos outros regimes as diferenças são mais difíceis de valorizar dado que se trata dum grupo pequeno (apenas constituído por 8 indivíduos). O tamanho do grupo condiciona também a amplitude do intervalo de confiança obtido (IC95%: 1,075 – 6,230). Há ainda a considerar a potencial importância duma avaliação a nível do fármaco, tal como realizado por Vo *et al.* [34]. No caso do presente trabalho tal não foi possível dada a grande variedade de esquemas utilizados [11]. O elevado número de esquemas obrigaria à estimação de um elevado número de coeficientes, o que não seria afetaria a qualidade do modelo, considerando o tamanho da amostra e a existência de outras variáveis.

A experiência terapêutica prévia com regimes não HAART foi incluída no modelo final mas não foi associada de forma estatisticamente significativa à primeira mudança terapêutica. Embora já não seja prática clínica a utilização de esquemas não HAART é importante salientar que nesta coorte ainda se encontram alguns desses doentes. Este grupo, tendo iniciado terapêutica num período com menos opções farmacológicas, já expostos a maior número de fármacos e possivelmente com experiências de falência prévia poderá também ter características próprias que, contudo, não se refletiram neste estudo.

Neste trabalho não foi possível incluir a variável habitualmente estudada como marcador de gravidade – a variável SIDA (existência ou não de imunodeficiência) – pelo elevado número de dados omissos [3]. Contudo, tanto a existência de neoplasias prévias à introdução de terapêutica HAART como a de infeções oportunistas traduzem a gravidade da infeção, previamente ao início de terapêutica, uma vez que se encontram associadas a estádios mais avançados da mesma. Ainda que não tenham sido incluídas em trabalhos prévios, o facto de habitualmente não se encontrar uma associação entre a existência de imunodeficiência e a mudança terapêutica está de acordo com os resultados obtidos neste estudo, no qual não se encontrou associação entre a mudança terapêutica e a existência de neoplasias ou infeções oportunistas [34].

O aumento da contagem de linfócitos CD4⁺ não se associou de forma estatisticamente significativa à primeira mudança. De salientar que os resultados previamente obtidos são variáveis: Landier *et al.* [24] descreveram em todas as classes de CD4⁺ abaixo de

200 células/ μ l um maior risco de mudança relativamente aos que tinham valor superior a este. Já em outros estudos como o apresentado por Vo *et al.* [34] a contagem de células CD4⁺ apresenta uma relação não linear com a mudança, existindo um risco superior para os participantes com contagens iguais ou superiores a 350/ μ l células mas também para os que apresentavam contagens inferiores a 200 células/ μ l (comparativamente à classe 200 a 350 células/ μ l). Contudo, nos restantes estudos analisados mais em detalhe (apresentados na Tabela 2) a contagem de linfócitos CD4⁺ não foi associada de forma estatisticamente significativa à mudança terapêutica. Assim, o resultado descrito está de acordo com os previamente publicados.

Um resultado importante é o obtido em relação à carga vírica. Esta variável aparece frequentemente associada à mudança terapêutica, mas, importa ressaltar que os diferentes estudos a analisam de forma e em momentos diferentes. Mocroft *et al.* [16] e Golub *et al.* [32] incluem esta variável como uma covariável dependente do tempo, obtendo um resultado estatisticamente significativo. Este resultado é expectável uma vez que a mudança da terapêutica é muitas vezes motivada por questões de falência terapêutica que se traduzem mais frequentemente por um aumento neste parâmetro laboratorial [2, 11]. Ao contrário destes estudos, no trabalho em questão foi analisada a associação entre a carga vírica antes do início de terapêutica HAART e a sua mudança, não tendo sido obtidos resultados estatisticamente significativos. Estes resultados estão de acordo com os descritos por de la Torre *et al.* [12]. Assim, a carga vírica pode estar associada à mudança terapêutica apenas quando se consideram os valores prévios à mudança e não quando analisamos os valores antes do início de terapêutica HAART.

Para além dos resultados do modelo importa também analisar a qualidade de ajustamento do mesmo. O R^2 foi calculado em 7,5%, um valor relativamente baixo, embora seja pouco frequente a obtenção de valores acima de 35% em modelos deste tipo [43]. Assim, poderão existir outras variáveis importantes na explicação da variável em estudo que não foram passíveis de ser incluídas no presente modelo. Também a probabilidade de concordância, com um valor de 60,1% (valor habitual para modelos de sobrevivência) aponta neste sentido [43]. Da restante avaliação não parecem ser violados os pressupostos do modelo.

Modelo para eventos múltiplos

As covariáveis selecionadas para constituírem o modelo final considerando as mudanças até à quarta são diferentes daquelas incluídas no modelo para a primeira mudança. A variável Período foi novamente incluída em todos os modelos, mesmo não apresentando significância estatística, para controlar pela época em que foi iniciada a terapêutica. No modelo final apenas a contagem de linfócitos $CD4^+$ foi associada de forma estatisticamente significativa à variável resposta. Assim, o aumento de 10 células/ μ l, que para valores baixos pode ter impacto clínico, está associado à diminuição de cerca de 1% no risco de mudança. Já um aumento de 100 células/ μ l, que será mais significativo para valores mais elevados, implicará uma diminuição de risco de mudança de cerca de 10%. Estes resultados parecem indicar que os doentes com um melhor estado imunológico se encontram em menor risco de mudança. A justificação destes resultados pode relacionar-se com características da própria infeção e com características do indivíduo. Por um lado, a infeção em estágio mais avançado poderá ter menor resposta às terapêuticas instituídas, com necessidade mais frequente de mudança. Por outro, o próprio indivíduo, diagnosticado numa fase mais precoce poderá ter um comportamento mais favorável perante a terapêutica – nomeadamente em questões de adesão.

Quando analisamos o gráfico de risco acumulado para cada uma das mudanças, verifica-se que existem diferenças entre a primeira mudança e as restantes. Embora num período inicial o risco de ocorrer uma primeira mudança siga um padrão semelhante ao das restantes, a partir do mês 35 há uma desaceleração no risco de primeira mudança. Este resultado parece sugerir que a primeira terapêutica HAART tem vindo a tornar-se mais estável, enquanto as restantes têm mantido o mesmo padrão. Estas diferenças verificam-se mesmo na presença da variável Período.

O facto de não serem conhecidos estudos prévios que tenham utilizado este tipo de metodologia dificulta a comparação com outras fontes bibliográficas. Por isso será necessário perceber se estes resultados se mantêm em futuras análises.

Limitações do estudo

O trabalho em discussão apresenta algumas limitações relacionadas com o estudo subjacente a estes dados mas também com a escolha dos modelos utilizados.

As limitações do estudo foram já descritas em detalhe previamente [11]. Salientam-se aqui as que mais diretamente influenciam a qualidade dos modelos obtidos.

Dado tratar-se de um estudo de coorte retrospectiva existem dados omissos para algumas variáveis que poderiam ser importantes para o modelo em estudo e não puderam ser incluídas (nomeadamente características sociodemográficas dos indivíduos em estudo como a naturalidade, nacionalidade, etnia ou situação profissional).

Na área da infeção VIH/SIDA são bem conhecidos os avanços na terapêutica, não só com a introdução da terapêutica HAART em 1996 como também a introdução sucessiva de vários fármacos e associações ao longo dos anos [3]. No presente trabalho a lista que serviu de base à seleção da amostra (aleatória simples) implicou a escolha de participantes que começaram a ser seguidos em diferentes momentos e, logo, com diferentes padrões terapêuticos. Neste sentido, a criação da variável Período é uma tentativa de evitar a influência deste fenómeno no modelo obtido: uma vez que esta variável foi sempre mantida na seleção de covariáveis, os resultados obtidos em todos os modelos estão controlados pelo período em que o doente iniciou terapêutica. Ainda assim, não é possível garantir que se esgotem as diferenças existentes nos três intervalos de tempo que foram considerados (com base em momentos em que ocorreram eventos chave na evolução do tratamento).

Na modelação do tempo até à mudança terapêutica e, como já referido anteriormente, todas as mudanças foram consideradas como representando exatamente o mesmo acontecimento. Contudo, e uma vez que as mudanças ocorrem por motivos diferentes é possível que este pressuposto não corresponda totalmente à realidade. Em alguns estudos anteriores, as mudanças foram divididas de acordo com o motivo pelo qual ocorrem, sendo esses acontecimentos modelados de forma separada [14, 22, 25, 33]. Dada a dimensão da amostra em estudo ($n = 194$) tal abordagem seria difícil neste trabalho, mesmo para a primeira mudança, e inviabilizaria o estudo das mudanças seguintes.

Alternativas à construção do modelo

Embora tenhamos optado pelo modelo PWP com os pressupostos anteriormente identificados, outras alternativas seriam possíveis. Essas alternativas incluem diferenças

quer na definição do evento de interesse quer na escolha dos modelos a serem implementados. O evento estudado, tal como justificado anteriormente, foi a mudança de pelo menos um dos fármacos do regime terapêutico, censurando os indivíduos aquando de uma paragem. Contudo, algumas abordagens alternativas a este nível poderiam ter sido consideradas:

- Considerar a paragem como uma censura intervalar, admitindo que enquanto não se encontravam sob terapêutica os doentes não estariam em risco de mudar. Esta estratégia para além de introduzir um esforço extra na construção da base de dados tornaria mais difícil a interpretação do modelo final e poderia introduzir um viés na estimação do risco, uma vez que uma paragem prévia poderia influenciar uma mudança futura.
- Considerar qualquer alteração na terapêutica (mudança de fármaco (M), paragem (P) ou reintrodução (R)) e utilizar os modelos multiestado, analisando as várias transições possíveis ($M \rightarrow P$, $P \rightarrow R$, $M \rightarrow M$, $R \rightarrow M$, $R \rightarrow P$). Esta estratégia embora aliciante responderia a uma pergunta de investigação diferente, mais abrangente, permitindo o cálculo da probabilidade de transição entre estados e covariáveis associadas a cada transição [43]. O número de indivíduos incluídos bem como as várias transições possíveis dificultariam a aplicação desta metodologia.
- Ao invés de ter sido aplicado um modelo de eventos múltiplos poder-se-ia ter optado por um modelo de fragilidade considerando um efeito aleatório de grupo, uma vez que vários indivíduos têm mais que uma mudança [43]. A inclusão de efeitos aleatórios na análise de dados relativos a mudança terapêutica foi previamente utilizada, embora com recurso à regressão de Poisson [35]. Ainda que não tenha sido desenvolvida no presente trabalho, esta será uma metodologia a considerar em trabalhos futuros.

Reflexão relativamente aos modelos de eventos múltiplos

A utilização de modelos de eventos múltiplos implica um bom conhecimento do tema em estudo para uma adequada seleção do modelo a utilizar. De referir que a construção da base de dados é um momento crucial e, em todo o processo, representa a maior parcela de dispêndio de tempo. Neste trabalho, o tempo dedicado à adequação da base de dados existente aos fins da análise, foi particularmente significativo. Isto deve-se ao

facto de a base de dados existente estar organizada em Consultas e não em eventos como era necessário para esta análise.

Uma nota final para a literatura existente em relação a estes modelos. Da pesquisa efetuada verifica-se que a literatura não é muito vasta e tem o seu pilar principal no livro de Terry Therneau e Patricia Grambsch – *Modeling Survival Data – Extending the Cox Model* [38]. Mesmo a implementação no programa utilizado nem sempre se encontra explícita, com necessidade de recorrer ao fórum de ajuda do programa para esclarecimento de algumas dúvidas.

Desta forma, é necessário continuar o desenvolvimento e adequação dos modelos existentes às necessidades da investigação médica, nomeadamente explorando questões como seleção de covariáveis que assume um papel muito relevante neste tipo de trabalhos.

Conclusão

O presente estudo permitiu concluir que os fatores associados à primeira mudança terapêutica diferem dos associados às restantes mudanças. Relativamente à primeira encontrou-se uma associação com o regime terapêutico inicial, quando se controla pelo período de início de terapêutica HAART e a experiência com terapêutica pré-HAART. Já quando se consideram as restantes mudanças apenas a contagem de células CD4⁺, controlando pelo período, se associa às mudanças terapêuticas.

Quando analisamos o risco acumulado de mudança terapêutica verifica-se que o risco de uma primeira seguiu uma evolução diferente das restantes mudanças, o que sugere uma alteração no padrão ao longo do tempo de estudo, provavelmente relacionada com a evolução terapêutica.

Relativamente à utilização de modelos de eventos múltiplos verifica-se a necessidade de selecionar o que melhor responde às questões em estudo, construindo a base a analisar em conformidade com o modelo escolhido.

Ainda que com algumas limitações os modelos utilizados aplicam-se a uma panóplia de questões de investigação na área médica e são mais um exemplo da transversalidade e importância da Bioestatística na investigação clínica.

Agradecimentos

Em primeiro lugar um agradecimento à minha orientadora – Professora Valeska Andreozzi, pelas suas críticas, sugestões e disponibilidade para esclarecimento de todas as minhas questões. Aproveito a oportunidade para estender o agradecimento a todos os elementos do Departamento de Estatística e Investigação Operacional que contribuíram para a minha formação neste mestrado, cujo presente trabalho representa o término.

Ao Professor Therry Therneau um merecido agradecimento, dado que sempre respondeu prontamente às questões que coloquei no fórum de ajuda do R e, como profundo conhecedor dos modelos aqui utilizados e autor da biblioteca do programa que serve de base a toda esta análise, sabiamente me aconselhou.

Como parte do projeto ATAR-VIH gostaria de agradecer ao Instituto de Medicina Preventiva pela oportunidade que me deu, a todos os elementos da equipa de investigação e às pessoas que contribuíram para a recolha e construção da base de dados (José Alexandre Freitas, Rita Martins, Maria Emília Pereira, Violeta Nogueira, João Costa). A um elemento em particular, Mestre Milene Fernandes, dirijo um agradecimento especial – por me ter permitido fazer do meu o meu trabalho e por todo o apoio nesta caminhada, nomeadamente como minha supervisora neste trabalho. Ainda ao Serviço de Doenças Infeciosas do Hospital de Santa Maria, na pessoa do seu diretor (Prof. Francisco Antunes) pela colaboração na recolha de dados e à Fundação Merck Sharp and Domme pelo seu apoio financeiro.

Aos meus amigos e familiares, um breve mas sentido agradecimento, por me terem dado a força necessária nos momentos mais difíceis.

Bibliografia

- [1] Likatavicius G, Van de Laar MJ. **HIV and AIDS in the European Union, 2009.** *Euro surveill* 2010, **15(48)**: 1-4.
- [2] Antunes F, Ed. **Manual sobre SIDA.** 2008, 3ª ed. Lisboa, Permanyer Portugal.
- [3] Fernandes M, **Adesão à Terapêutica Anti-Retroviral em indivíduos seropositivos para o VIH-1: prevalência e factores associados.** Tese de mestrado, Faculdade de Medicina da Universidade de Lisboa, 2011.
- [4] Fauci A, **HIV and AIDS: 20 years of science - Therapy for HIV infection.** *Nat Med* 2003, **9(7)**: 839-43.
- [5] Coordenação Nacional para a Infecção VIH/SIDA, **Recomendações Portuguesas para o Tratamento da Infecção VIH/SIDA.** 2011. Disponível em <http://sm.vectweb.pt/media/64/File/Actualidades/portugal/Tratamento%20VIH%20em%20Portugal%20-%202011%20CNSIDA.pdf>, acedido a 18-10-2012.
- [6] O'Brien ME, Clark RA, Besch CL, Myers L, Kissinger P. **Patterns and correlates of discontinuation of the initial HAART regimen in an urban outpatient cohort.** *J Acquir Immune Defic Syndr* 2003, **34(4)**: 407-14.
- [7] Kirstein LM *et al.* **Prevalence and Correlates of Highly Active Antiretroviral Therapy Switching in the Women's Interagency HIV Study.** *J Acquir Immune Defic Syndr* 2002, **29(5)**: 495-503.
- [8] Grupo de Estudio VIH/VIR. **Estudio epidemiológico retrospectivo sobre la duración del tratamiento de la infección por el virus de la inmunodeficiencia humana en España.** *Med Clin (Barc)* 2002, **119(19)**: 721-4.
- [9] Nogueira J, Reis J. **Terapêutica de doentes VIH em ambulatório no Hospital Militar de Belém - estudo retrospectivo.** *V Congresso Nacional da Associação Portuguesa dos Farmacêuticos Hospitalares*, 2004 [resumo nº 35]. Disponível em <http://www.apfh.pt/scid/webApfh/defaultArticleViewOne.asp?categoryID=667&articleID=690>, acedido a 18-10-2012.
- [10] Caetano S *et al.* **Motivos para a alteração da terapêutica anti-retroviral dos doentes com infecção VIH-1.** *X Congresso de Doenças Infecciosas e Microbiologia Clínica/ VIII Congresso Nacional sobre SIDA*, 2010 [resumo nº PO116].
- [11] Leite A. **Frequency and reasons for antiretroviral therapy switching in HIV patients.** Tese de Mestrado, Faculdade de Medicina da Universidade de Lisboa, 2011.
- [12] De la Torre J *et al.* **Primera pauta de tratamiento antiretroviral en pacientes con infección por el virus de la inmunodeficiencia humana: durabilidad y factores asociados a su modificación.** *Enferm Infecc Microbiol Clin* 2008, **26(7)**: 416-22.

- [13] van Roon EN, Verzijl JM, Juttmann JR, Lenderink AW, Blans MJ, Egberts ACG. **Incidence of Discontinuation of Highly Active Antiretroviral Combination Therapy (HAART) and Its Determinants.** *JAIDS* 1999, **20(3)**: 290-4.
- [14] d' Arminio Monforte *et al.* **Insights into the reasons for discontinuation of the first highly active antiretroviral therapy (HAART) regimen in a cohort of antiretroviral naïve patients.** *AIDS* 2000, **14(5)**: 499-507.
- [15] Dorruci M *et al.* **Time to discontinuation of the first highly active antiretroviral therapy regimen: a comparison between protease inhibitor- and non-nucleoside reverse transcriptase inhibitor-containing regimens.** *AIDS* 2001, **15(13)**: 1733-6.
- [16] Mocroft A *et al.* **Reasons for modification and discontinuation of antiretrovirals : results from a single treatment centre.** *AIDS* 2001, **15(2)**: 185-94.
- [17] Dieleman JP *et al.* **Determinants of recurrent toxicity-driven switches of highly active antiretroviral therapy.** *AIDS* 2002, **16(5)**: 737-45.
- [18] Ahdieh-grant *et al.* **Factors and Temporal Trends Associated With Highly Active Antiretroviral Therapy Discontinuation in the Women's Interagency HIV Study.** *J Acquir Immune Defic Syndr* 2005, **38(4)**: 500-3.
- [19] Li X *et al.* **Interruption and discontinuation of highly active antiretroviral therapy in the multicenter AIDS cohort study.** *J Acquir Immune Defic Syndr* 2005, **38(3)**: 320-8.
- [20] Yuan Y, L'Italien G, Mukherjee J, Iloeje UH. **Determinants of discontinuation of initial highly active antiretroviral therapy regimens in a US HIV-infected patient cohort.** *HIV Med* 2006, **7(3)**: 156-62.
- [21] Sabin CA *et al.* **The associations between age and the development of laboratory abnormalities and treatment discontinuation for reasons other than virological failure in the first year of highly active antiretroviral therapy.** *HIV Med* 2009, **10(1)**: 35-43.
- [22] Cardoso SW *et al.* **Incidence of modifying or discontinuing first HAART regimen and its determinants in a cohort of HIV-infected patients from Rio de Janeiro, Brazil.** *AIDS Res Hum Retroviruses* 2010, **26(8)**: 865-74.
- [23] Cesar C *et al.* **Rates and reasons for early change of first HAART in HIV-1-infected patients in 7 sites throughout the Caribbean and Latin America.** *PloS One* 2010, **5(6)**: e10490.
- [24] Landier J *et al.* **Switch to second-line ART in West African routine care: incidence and reasons for switching.** *AIDS Care* 2011, **23(1)**: 75-8.
- [25] Hänsel A, Bucher H, Nüesch R, Battegay M. **Reasons for discontinuation of first highly active antiretroviral therapy in a cohort of proteinase inhibitor-naïve HIV-infected patients.** *J Acquir Immune Defic Syndr* 2001, **26(2)**: 191-3.
- [26] Kerr T *et al.* **Determinants of HAART discontinuation among injection drug users.** *AIDS Care* 2005 **17(5)**: 539-49.

- [27] Kumarasamy N *et al.* **Reasons for modification of generic highly active antiretroviral therapeutic regimens among patients in southern India.** *J Acquir Immune Defic Syndr* 2006, **41(1)**: 53-8.
- [28] Kiguba R, Byakika-Tusiime J, Karamagi C, Ssali F, Mugenyi P, Katabira E. **Discontinuation and modification of highly active antiretroviral therapy in HIV-infected Ugandans: prevalence and associated factors.** *J Acquir Immune Defic Syndr* 2007, **45(2)**: 218-23.
- [29] Maggiolo F *et al.* **Reasons for discontinuation of nevirapine-containing HAART: results from an unselected population of a large clinical cohort.** *J Antimicrob Chemother* 2007, **59(3)**: 569-72.
- [30] Himelhoch S *et al.* **HIV patients with psychiatric disorders are less likely to discontinue HAART.** *AIDS* 2009, **23(13)**: 1735-42.
- [31] Mirjam-Colette K, Pisu M, Dumcheva A, Westfall O, Kilby M, Saag MS. **Gender differences in discontinuation of antiretroviral treatment regimens.** *J Acquir Immune Defic Syndr* 2010, **52(3)**: 336-41.
- [32] Golub ET *et al.* **Patterns, predictors, and consequences of initial regimen type among HIV-infected women receiving highly active antiretroviral therapy.** *Clin Infect Dis* 2008, **46(2)**: 305-12.
- [33] Robison LS *et al.* **Short-term discontinuation of HAART regimens more common in vulnerable patient populations.** *AIDS Res Hum Retroviruses* 2008, **24(11)**: 1347-55.
- [34] Vo TTN *et al.* **Durability and outcome of initial antiretroviral treatments received during 2000-2005 by patients in the Swiss HIV Cohort Study.** *J Infect Dis* 2008, **197(12)**: 1685-94.
- [35] Srasuebkul P, Calmy A, Zhou J, Kumarasamy N, Law M, Lim PL. **Impact of drug classes and treatment availability on the rate of antiretroviral treatment change in the TREAT Asia HIV Observational Database (TAHOD).** *AIDS Res Ther* 2007, **4**: 18.
- [36] Ripamonti D, Arici C, Pezzotti P, Maggiolo F, Ravasio L, Suter F. **Hepatitis C infection increases the risk of the modification of first highly active antiretroviral therapy in HIV-infected patients.** *AIDS* 2004, **18(2)**: 334-7.
- [37] R Development Core Team. **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0. Disponível em <http://www.R-project.org/>, acessado a 18-10-2012
- [38] Therneau T, Grambsch P. **Modelling Survival Data: Extendign the Cox Model.** 1st ed. New York, United States: Springer, 2000.
- [39] Therneau T. **A Package for Survival Analysis in R.** R package version 2.36-12. 2012.
- [40] Kelly PJ, Lim LLY. **Survival analysis for recurrent event data: an application to childhood infectious diseases.** *Statist Med* 2000, **19(1)**: 13-33.

- [41] Wei LJ, Glidden DV. **An overview of statistical methods for multiple failure time data in clinical trials.** *Statist Med* 1997, **16(8)**: 833-9.
- [42] Leclerc BS *et al.* **Risk factors for falling among community-dwelling seniors using home-care services: an extended hazards model with time-dependent covariates and multiple events.** *Chronic Dis Can* 2008, **28(4)**: 111-20.
- [43] Carvalho M, Andreozzi V, Codeço C, Campos D, Barbosa M, Shimakura S. **Análise de Sobrevida - Teoria e aplicações em Saúde.** 2011, 2ª ed, Rio de Janeiro. Editora Fiocruz.
- [44] Castañeda J, Gerritse B. **Appraisal of Several Methods to Model Time to Multiple Events per Subject: Modelling Time to Hospitalizations and Death.** *Revista Colombiana de Estadística* 2010, **33(1)**: 43-61.
- [45] Carvalho M, Andreozzi V, Codeço C, Campos D, Barbosa M, Shimakura S. **Análise de sobrevida - Teoria e aplicações em saúde.** 2005, 1ª ed, Rio de Janeiro. Editora Fiocruz.
- [46] Prentice R, Williams B, Peterson A. **On the regression analysis of multivariate failure time data.** *Biometrika* 1981, **68(2)**: 373-9.
- [47] Turkman MA, Silva G. **Modelos Lineares Generalizados - da teoria à prática.** 2000, Sociedade Portuguesa de Estatística, 94-8.
- [48] Instituto Nacional de Saúde Dr. Ricardo Jorge. **Infecção VIH/SIDA - A situação em Portugal a 31 de Dezembro de 2009.** 2010.
- [49] Gratacòs L *et al.* **Tratamiento antirretroviral de la infección por el virus de la inmunodeficiencia humana: duración y motivos de cambio del primer esquema terapéutico en 518 pacientes.** *Med Clin (Barc)* 2006, **126(7)**: 241-5.

Anexos

Construção da base de dados a analisar

```
#Base de consultas com variáveis de interesse
Consultas2<-
Consultas[,c("idPARTICIPANTE", "TipoTAR", "MudancaQual", "MudancaMo
tivo",
"Ef Secundário - Qual", "Outro - qual", "Motivo Combinado - Qual",
"Sem TAR - motivo", "Co-morbilidade-
qual", "MudancaNotas", "DataConsulta",
"CD4num", "CV", "CVnum")]

#Linhas primeira consulta HAART de cada indivíduo
y<-which(Consultas$TipoTAR=="HAART")#Posição das consultas HAART
emHAART<-c(y[1])
for(i in 2:length(y)){
  if (Consultas$idPARTICIPANTE[y[i-
1]]!=Consultas$idPARTICIPANTE[y[i]]){
    emHAART<-c(emHAART,y[i])
  }
}

#Linhas da última consulta de cada indivíduo
ultHAART<-c(linhaUltima,dim(Consultas)[1])

#Colocar HAART em todas as consultas entre a primeira e última
Consultas2$HAART<-NA
for(i in 1:length(emHAART)){
  Consultas2$HAART[emHAART[i]:ultHAART[i]]<-"HAART"}

#Começar a ver mudanças
Mudancas<-Consultas2[which((Consultas2$MudancaQual=="Farmaco" |
Consultas2$MudancaQual=="Sem
TAR"|Consultas2$MudancaQual=="Reinicio e Farmaco"|
Consultas2$MudancaQual=="Reinicio de
TAR")&(Consultas2$HAART=="HAART")),c("idPARTICIPANTE", "TipoTAR",
"MudancaQual", "MudancaMotivo",
"Ef Secundário - Qual", "Outro - qual", "Motivo Combinado - Qual",
"Sem TAR - motivo", "Co-morbilidade-
qual", "MudancaNotas", "DataConsulta",
"CD4num", "CV", "CVnum")]

#Datas - função para obter a data de início de HAART de cada
indivíduo
Datas<-function(x){
```

```

datasTAR$dataHAART[which(datasTAR$idPARTICIPANTE==x)]}
#criar passo intermédio para tirar as mudanças que são início de
HAART
guardar<-c()
for(i in 1:length(Mudancas$idPARTICIPANTE)){
  if((Datas(Mudancas$idPARTICIPANTE[i])-Mudancas$data[i]!=0))
  guardar<-c(guardar,i)}

#Mudanças a considerar
MudancasN<-Mudancas[guardar,]

#Considerar apenas mudança de fármaco
#Criar uma coluna para ver se é para manter (Fármaco) ou para
censurar
mudB<-c()
for(i in 1:dim(MudancasN)[1]){
  if(MudancasN$MudancaQual[i]!="Farmaco")
  mudB<-c(mudB,"Censurar")
  else
  mudB<-c(mudB,"Farmaco")}
MudancasN$mudB<-mudB

MudTirarC<-MudancasN
for (i in 1:dim(MudTirarC)[1]){
  if(MudTirarC[i,18]=="Censurar"){
    id<-MudTirarC[i,1]
    ver<-which(match(MudTirarC$idPARTICIPANTE,id)!="NA")
    for(p in 1:length(ver)){
      if(ver[p]>i&&MudTirarC[ver[p],18]=="Farmaco"){
        MudTirarC[ver[p],18]<-"Censurar"
      }
    }
  }
}

#Número de mudanças por indivíduo
totalMudF<-
table(MudTirarC$idPARTICIPANTE[(MudTirarC$mudB=="Farmaco")])
#ver quem não tem mudanças
posSemMud<-
which(is.na(match(unique(Consultas2$idPARTICIPANTE),row.names(to
talMudF))))
num[posSemMud]

#Resumir todas as mudnaças
MTotal<-table(t(t(totalMudF))[,1])

#Criar data frame só com as mudanças a considerar
MudFarmaco<-MudTirarC[which(MudTirarC$mudB=="Farmaco"),]

```

```

aux1<-c(1)
for(i in 2:length(MudFarmaco$idPARTICIPANTE)){
  if(MudFarmaco$idPARTICIPANTE[i-
1]!=MudFarmaco$idPARTICIPANTE[i]){
    aux1<-c(aux1,i)
  }

#####
#Primeira Mudança
#####

#dataframe das primeiras mudanças
primMud<-MudFarmaco[aux1,]
vectorDatas<-c()
for (i in primMud$idPARTICIPANTE){
  vectorDatas<-c(Datas(i),vectorDatas)
}
vctorDatas<-as.Date(vectorDatas,yyyy-mm-dd)
vDatas<-c()
for(i in 1:length(vctorDatas)){
  vDatas<-c(vctorDatas[i],vDatas)
}

#Tempo até primeira mudança
datasTAR$dataHAART<-as.Date(datasTAR$dataHAART,format="%Y-%m-
%d")
primMud$vDatas<-vDatas
primMud$data<-as.Date(primMud$dataConsulta,format="%Y-%m-%d")
primMud$tempo<-as.numeric((primMud$data-primMud$vDatas)/30)
KM<-
data.frame(datasTAR$idPARTICIPANTE,datasTAR$dataHAART,datasTAR$d
ataUltima)
colnames(KM)<-c("idPARTICIPANTE","dataHAART","dataUltima")

#Confirmar os que são para censurar ou não
m<-match(KM$idPARTICIPANTE,primMud$idPARTICIPANTE)
KM$status<-NA
for (i in 1:length(KM$idPARTICIPANTE)){
  if(is.na(m[i])){
    KM$status[i]<-0
  }else{
    KM$status[i]<-1
  }
}

#Consultas paragem - corrigir data
auxPar<-Consultas[which(Consultas$MudancaQual=="Sem TAR"),1:3]

```

```

auxPar[,3]<-NA
auxPar[,4]<-NA
for(i in 1:dim(auxPar)[1]){
  auxPar[i,3]<-
  as.character(KM$dataHAART[which(KM$idPARTICIPANTE==auxPar
  [i,1])])}

for(i in 1:dim(auxPar)[1]){
  auxPar[i,4]<-
  as.character(KM$dataUltima[which(KM$idPARTICIPANTE==auxPar
  [i,1])])}
colnames(auxPar)[2]<-"DataParagem"
colnames(auxPar)[3]<-"DataHAART"
colnames(auxPar)[4]<-"DataUltima"

#Tirar os que têm paragens antes da HAART
auxPar$difHAARTP<-as.Date(auxPar[,2],format="%Y-%m-%d") -
as.Date(auxPar[,3],format="%Y-%m-%d")
auxPar<- auxPar[-which(auxPar[,5]<0),]
#Mudar a data da última
dataUltimaSubs<-c(as.character(auxPar[1,2]))
for(i in 2:dim(auxPar)[1]){
  if(auxPar[i,1]!= auxPar[i-1,1])
  dataUltimaSubs<-c(dataUltimaSubs,as.character(auxPar[i,2]))}
length(unique(auxPar[,1]))
length(dataUltimaSubs)

KM$dataUltima[which(is.element(KM$idPARTICIPANTE,unique(auxPar[,
1])))]<-
as.Date(dataUltimaSubs,format="%Y-%m-%d")

#Calcular início e fim
hoje<-Sys.Date()
dif<-c()
for(i in 1:dim(KM)[1]){
  dif<-c(dif,hoje-KM$dataHAART[i])}
posRef<-which(dif==max(dif))
KM$ini<-as.numeric((KM$dataHAART-KM$dataHAART[posRef])/30)
KM$fim<-NA
for(i in 1:dim(KM)[1]){
  if(KM$status[i]==0)
  KM$fim[i]<-as.numeric((KM$dataUltima[i]-
  KM$dataHAART[posRef])/30)}

#Colocar as datas da primMud na dataframe KM
posMudKM<-which(is.na(KM$fim))#posição dos que têm status 1
for(p in 1:length(posMudKM)){
  KM$fim[posMudKM[p]]<-as.numeric((primMud$data[p]-

```

```

KM$dataHAART[posRef])/30)}
KM$tempo<-KM$fim-KM$ini

#Restantes covariáveis obtidas a partir da base de dados inicial

#####
#Segunda Mudança
#####

#dataFrame das segundas mudanças
aux2a<-c()
for(i in 1:length(totalMudF)){
  if(totalMudF[[i]]>1)
  aux2a<-c(aux2a,aux1[i]+1)
}
segMud<-MudFarmaco[aux2a,]

#Linhas a considerar na KM e primMud
#KM
linhasSegMud<-
which(match(KM$idPARTICIPANTE,segMud$idPARTICIPANTE)!="NA")
#primMud
linhasPrimSegMud<-
which(match(primMud$idPARTICIPANTE,segMud$idPARTICIPANTE)!="NA")

#linhasSegMud - posições que são utilizadas no KM para comparar
com o KM2
#linhasPrimSegMud - posições que são utilizadas no KM para
comparar com o primMud

#ini=KM$fim
#fim=data da 2ª mudança-data da primeira mudança+ini
KM2<-as.data.frame(segMud$idPARTICIPANTE)
colnames(KM2)[1]<-"idPARTICIPANTE"
KM2$dataHAART<-KM$dataHAART[linhasSegMud]
KM2$dataUltima<-KM$dataUltima[linhasSegMud]
KM2$status<-1#todos neste grupo tem uma mudança
KM2$ini<-KM$fim[linhasSegMud]
KM2$fim<-(as.numeric(segMud$data-
primMud$data[linhasPrimSegMud])/30)+KM2$ini
KM2$tempo<-as.numeric(segMud$data-
primMud$data[linhasPrimSegMud])/30
KM2$Sexo<-KM$Sexo[linhasSegMud]

for(i in 10:19){
  KM2[,i]<-KM[linhasSegMud,i]}
colnames(KM2)[10:19]<-colnames(KM)[10:19]

```

```
#####
#Terceira Mudança
#####

#dataFrame das terceiras mudanças
aux3<-c()
for(i in 1:length(totalMudF)){
  if(totalMudF[[i]]>2)
    aux3<-c(aux3,aux1[i]+2)
}
terMud<-MudFarmaco[aux3,]
dim(terMud)

#KM2
linhasTerMud<-
which(match(KM2$idPARTICIPANTE,terMud$idPARTICIPANTE)!="NA")
#segMud
linhasSegTerMud<-
which(match(segMud$idPARTICIPANTE,terMud$idPARTICIPANTE)!="NA")

#linhasTerMud - posições que são utilizadas no KM para comparar
com o KM2
#linhasSegTerMud - posições que são utilizadas no KM para
comparar com o segMud

#ini=KM2$fim
#fim=data da 3ª mudança-data da 2ªmudança+KM2$ini
KM3<-as.data.frame(terMud$idPARTICIPANTE)
colnames(KM3)[1]<-"idPARTICIPANTE"
KM3$dataHAART<-KM2$dataHAART[linhasTerMud]
KM3$dataUltima<-KM2$dataUltima[linhasTerMud]
KM3$status<-1#todos neste grupo tem uma mudança
KM3$ini<-KM2$fim[linhasTerMud]
KM3$fim<-(as.numeric(terMud$data-
segMud$data[linhasSegTerMud])/30)+KM3$ini
KM3$tempo<-as.numeric(terMud$data-
segMud$data[linhasSegTerMud])/30
KM3$Sexo<-KM2$Sexo[linhasTerMud]

for(i in 10:19){
  KM3[,i]<-KM2[linhasTerMud,i]}
colnames(KM3)[10:19]<-colnames(KM)[10:19]

#####
#Quarta Mudança
#####

#dataFrame das quartas mudanças
```



```

aux4<-c()
for(i in 1:length(totalMudF)){
  if(totalMudF[[i]]>3)
    aux4<-c(aux4,aux1[i]+3)
}
quarMud<-MudFarmaco[aux4,]
dim(quarMud)

#KM3
linhasQuarMud<-
which(match(KM3$idPARTICIPANTE,quarMud$idPARTICIPANTE)!="NA")
#terMud
linhasTerQuarMud<-
which(match(terMud$idPARTICIPANTE,quarMud$idPARTICIPANTE)!="NA")

#linhasQuarMud - posições que são utilizadas no KM para comparar
com o KM2
#linhasTerQuarMud - posições que são utilizadas no KM para
comparar com o segMud

#ini=KM3$fim
#fim=data da 4ª mudança-data da 3ªmudança+KM3$ini
KM4<-as.data.frame(quarMud$idPARTICIPANTE)
colnames(KM4)[1]<-"idPARTICIPANTE"
KM4$dataHAART<-KM3$dataHAART[linhasQuarMud]
KM4$dataUltima<-KM3$dataUltima[linhasQuarMud]
KM4$status<-1#todos neste grupo tem uma mudança
KM4$ini<-KM3$fim[linhasQuarMud]
KM4$fim<-(as.numeric(quarMud$data-
terMud$data[linhasTerQuarMud])/30)+KM4$ini
KM4$tempo<-as.numeric(quarMud$data-
terMud$data[linhasTerQuarMud])/30
KM4$Sexo<-KM3$Sexo[linhasQuarMud]

for(i in 10:19){
  KM4[,i]<-KM3[linhasQuarMud,i]}
colnames(KM4)[10:19]<-colnames(KM)[10:19]

#####
#Sem mudanças
#####

#####
#Primeira Mudança
#####
#Para os que só têm uma mudança

```

```

idPrim<-
KM$idPARTICIPANTE[which(KM$status==1 & !(KM$idPARTICIPANTE%in%KM2
$idPARTICIPANTE)))]

#Retirar da primMud e KM as linhas equivalentes
#KM
linhasFinal1KM<-which(match(KM$idPARTICIPANTE,idPrim)!="NA")
#primMud
linhasFinal1<-which(match(primMud$idPARTICIPANTE,idPrim)!="NA")

#linhasFinal1KM - posições que são utilizadas no KM para
comparar
#linhasFinal1 - posições que são utilizadas para comparar com o
Mud anterior

#ini=KM$fim
#fim=data da ultima consulta-data da 1ªmudança+KMsemMud$ini
KMsemMud<-as.data.frame(idPrim)
colnames(KMsemMud)[1]<-"idPARTICIPANTE"
KMsemMud$dataHAART<-KM$dataHAART[linhasFinal1KM]
KMsemMud$dataUltima<-KM$dataUltima[linhasFinal1KM]
KMsemMud$status<-0#acabam o seguimento
KMsemMud$ini<-KM$fim[linhasFinal1KM]

KMsemMud$fim<-(as.numeric(KM$dataUltima[linhasFinal1KM]-
primMud$data[linhasFinal1])/30)+KMsemMud$ini
KMsemMud$tempo<-as.numeric(KM$dataUltima[linhasFinal1KM]-
primMud$data[linhasFinal1])/30
KMsemMud$Sexo<-KM$Sexo[linhasFinal1KM]
KMsemMud$Transmissao<-KM$Transmissao[linhasFinal1KM]

for(i in 10:19){
KMsemMud[,i]<-KM[linhasFinal1KM,i]}
colnames(KMsemMud)[10:19]<-colnames(KM)[10:19]

#1º verificar que na consulta final não houve mudança
verUltimaData<-c()
for (i in 1:length(linhasFinal1)){
  if(primMud$data[linhasFinal1][i]==KM$dataUltima[linhasFinal
1KM][i])
    verUltimaData<-c(verUltimaData,i)
}

#retirar aqueles cuja última mudança corresponde a última
Mudança
KMsemMud<-KMsemMud[-verUltimaData,]

#criar o estrato

```

```

KM$Estrato<-1
KMSemMud$Estrato<-2

#Juntar as dataframes
KM1Conj<-rbind(KM[, -20], KMSemMud)
KM1Unido<-
KM1Conj[order(KM1Conj$idPARTICIPANTE, KM1Conj$Estrato), ]

#####
#Segunda Mudança
#####
#Para os que só têm duas mudança
idSeg<-
KM2$idPARTICIPANTE[which(KM2$status==1 & !(KM2$idPARTICIPANTE%in%
KM3$idPARTICIPANTE))]

#Retirar da segMud e KM2 as linhas equivalentes
#KM2
linhasFinal2KM<-which(match(KM2$idPARTICIPANTE, idSeg)!="NA")
#segMud
linhasFinal2<-which(match(segMud$idPARTICIPANTE, idSeg)!="NA")

#linhasFinal2KM - posições que são utilizadas no KM para
comparar
#linhasFinal2 - posições que são utilizadas para comparar com o
Mud anterior

#ini=KM2$fim
#fim=data da ultima consulta-data da 2ªmudança+KMSemMud2$ini (ou
seja o tempo que passou
#entre as 2 mudanças e adicionar o tempo que o doente já foi
seguido)
KMSemMud2<-as.data.frame(idSeg)
colnames(KMSemMud2)[1]<-"idPARTICIPANTE"
KMSemMud2$dataHAART<-KM2$dataHAART[linhasFinal2KM]
KMSemMud2$dataUltima<-KM2$dataUltima[linhasFinal2KM]
KMSemMud2$status<-0#acabam o seguimento
KMSemMud2$ini<-KM2$fim[linhasFinal2KM]
KMSemMud2$fim<-(as.numeric(KM2$dataUltima[linhasFinal2KM]-
segMud$data[linhasFinal2])/30)+KMSemMud2$ini
KMSemMud2$tempo<-as.numeric(KM2$dataUltima[linhasFinal2KM]-
segMud$data[linhasFinal2])/30
KMSemMud2$Sexo<-KM2$Sexo[linhasFinal2KM]
KMSemMud2$Transmissao<-KM2$Transmissao[linhasFinal2KM]

for(i in 10:19){
KMSemMud2[, i]<-KM2[linhasFinal2KM, i]}
colnames(KMSemMud2)[10:19]<-colnames(KM)[10:19]

```

```

#1º verificar que na consulta final não houve mudança
verUltimaData2<-c()
for (i in 1:length(linhasFinal2)){
  if(segMud$data[linhasFinal2][i]==KM2$dataUltima[linhasFinal
2KM][i])
    verUltimaData2<-c(verUltimaData2,i)
}

#retirar aqueles cuja ultima mudança corresponde à última
Mudança
KM2SemMud2<-KM2SemMud2[-verUltimaData2,]

#criar o estrato
KM2$Estrato<-2
KM2SemMud2$Estrato<-3
#começar por adicionar as ultimas linhas à df "mae",ordenar e
depois unir todas a ordenar
KM2Conj<-rbind(KM2,KM2SemMud2)
KM2Unido<-
KM2Conj[order(KM2Conj$idPARTICIPANTE,KM2Conj$Estrato),]

#####
#Terceira Mudança
#####
#Para os que só têm três mudança
idTer<-
KM3$idPARTICIPANTE[which(KM3$status==1&!(KM3$idPARTICIPANTE%in%
KM4$idPARTICIPANTE)))]

#Retirar da terMud e KM3 as linhas equivalentes
#KM3
linhasFinal3KM<-which(match(KM3$idPARTICIPANTE,idTer)!="NA")
#terMud
linhasFinal3<-which(match(terMud$idPARTICIPANTE,idTer)!="NA")

#linhasFinal3KM - posições que são utilizadas no KM para
comparar
#linhasFinal3 - posições que são utilizadas para comparar com o
Mud anterior

#ini=KM3$fim
#fim=data da ultima consulta-data da 3ªmudança+KM2SemMud3$ini (ou
seja o tempo que passou
#entre as 2 mudanças e adicionar o tempo que o doente já foi
seguido)
KM2SemMud3<-as.data.frame(idTer)
colnames(KM2SemMud3)[1]<- "idPARTICIPANTE"

```

```

KM3SemMud3$dataHAART<-KM3$dataHAART[linhasFinal3KM]
KM3SemMud3$dataUltima<-KM3$dataUltima[linhasFinal3KM]
KM3SemMud3$status<-0#acabam o seguimento
KM3SemMud3$ini<-KM3$fim[linhasFinal3KM]
KM3SemMud3$fim<-(as.numeric(KM3$dataUltima[linhasFinal3KM]-
terMud3$data[linhasFinal3])/30)+KM3SemMud3$ini
KM3SemMud3$tempo<-as.numeric(KM3$dataUltima[linhasFinal3KM]-
terMud3$data[linhasFinal3])/30
KM3SemMud3$Sexo<-KM3$Sexo[linhasFinal3KM]
KM3SemMud3$Transmissao<-KM3$Transmissao[linhasFinal3KM]

for(i in 10:19){
KM3SemMud3[,i]<-KM3[linhasFinal3KM,i]}
colnames(KM3SemMud3)[10:19]<-colnames(KM)[10:19]

#1º verificar que na consulta final nao houve mudança
verUltimaData3<-c()
for (i in 1:length(linhasFinal3)){
  if(terMud3$data[linhasFinal3][i]==KM3$dataUltima[linhasFinal
3KM][i])
    verUltimaData3<-c(verUltimaData3,i)
}

#retirar aqueles cuja ultima mudança corresponde à última
Mudança
KM3SemMud3<-KM3SemMud3[-verUltimaData3,]

#criar o estrato
KM3$Estrato<-3
KM3SemMud3$Estrato<-4

#Juntar as dataframes
KM3Conj<-rbind(KM3,KM3SemMud3)
KM3Unido<-
KM3Conj[order(KM3Conj$idPARTICIPANTE,KM3Conj$Estrato),]
KMTTotal<-rbind(KM1Unido,KM2Unido,KM3Unido,KM4)
KM4$Estrato<-4

#Dataframe final
PWP<-KMTTotal[order(KMTTotal$idPARTICIPANTE,KMTTotal$Estrato),]

```

Análise exploratória

```

####
#Sexo
####
MSexoC<-survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Sexo,
data = PrimMudanca)

```

```

MSexoC
plot(MSexoC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1,2))
legend(x="bottomleft",legend=c("Feminino","Masculino"),lty=c(1,2),
),title="Sexo",bty="n",cex=0.8)

#####
#Regime
#####
MRegimeC<-
survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Regime, data
= PrimMudanca)
MRegimeC
plot(MRegimeC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:5))
legend(x="topright",legend=c(levels(PrimMudanca$Regime)),lty=c(1
:5),title="Regime",bty="n",cex=0.8)

#####
#HAART
#####
MHAARTC<-
survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~HAART, data =
PrimMudanca)
MHAARTC
plot(MHAARTC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:2))
legend(x="topright",legend=c(levels(as.factor(PrimMudanca$HAART)
)),lty=c(1:5),title="HAART",bty="n",cex=0.8)

#####
#Drogas
#####
MDrogasC<-
survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Drogas, data
= PrimMudanca)
MDrogasC
plot(MDrogasC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:2))
legend(x="topright",legend=c(levels(as.factor(PrimMudanca$Drogas
))),lty=c(1:2),title="Consumo drogas",bty="n",cex=0.8)

#####
#Psi
#####
MPsiC<-survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Psi,
data = PrimMudanca)
MPsiC

```

```

plot(MPsiC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:2))
legend(x="topright",legend=c(levels(as.factor(PrimMudanca$Psi)))
,lty=c(1:2),title="Antecedentes psiquiátricos",bty="n",cex=0.8)

#####
#Neo
#####
MNeoC<-survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Neo,
data = PrimMudanca)
MNeoC
plot(MNeoC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:2))
legend(x="topright",legend=c(levels(as.factor(PrimMudanca$Neo)))
,lty=c(1:2),title="Antecedentes de neoplasia",bty="n",cex=0.8)

#####
#IO
#####
MIOC<-survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~IO,
data = PrimMudanca)
MIOC
plot(MIOC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:2))
legend(x="topright",legend=c(levels(as.factor(PrimMudanca$IO))),
lty=c(1:2),title="Antecedentes
infeções
oportunisticas",bty="n",cex=0.8)

#####
#Periodo
#####
MPeriodoC<-
survfit(Surv(PrimMudanca$tempo,PrimMudanca$status)~Periodo, data
= PrimMudanca)
MPeriodoC
plot(MPeriodoC,mark.time=F,conf.int=F, xlab="Meses",ylab="S(t)",
main="Função de sobrevivência",lty=c(1:3))
legend(x="topright",legend=c("Antes 2000","2000-2004","2005 em
diante"),
lty=c(1:3),title="Período de início de
terapêutica",bty="n",cex=0.8)

```

Seleção de covariáveis (Modelo de Cox)

```

deviance <- function(x) {
modelo
coxph(Surv(ini,fim,status)~Periodo+PrimMudanca[,x],data=PrimMudanca, x=T)

```

```

teste <- anova(modelo, test = "Chisq")
df <- teste$Df[3]
p <- teste[3, 4]
resp <- data.frame(df, p)
names(resp) <- c("Df", "Pvalue")
resp
}

mB<-seq(8,18)
selectCov<-data.frame(Variavel      =      names(PrimMudanca[mB]),
t(sapply(mB, deviance)))
selectCov[order(unlist(selectCov$Pvalue)),]

#Seleção co variaveis
deviance2 <- function(x) {
  modelo1                                     <-
coxph(Surv(ini,fim,status)~Periodo+Regime,data=PrimMudanca,x=T)
  modelo2                                     <-
coxph(Surv(ini,fim,status)~Periodo+Regime+PrimMudanca[,x],data=P
rimMudanca,x=T)
  teste <- anova(modelo1,modelo2,test = "Chisq")
  df <- teste$Df[2]
  p <- teste[2, 4]
  resp <- data.frame(df, p)
  names(resp) <- c("Df", "Pvalue")
  resp
}

mB2<-seq(8,18)[-3]
selectCov2<-data.frame(Variavel      =      names(PrimMudanca[mB2]),
t(sapply(mB2, deviance2)))
selectCov2[order(unlist(selectCov2$Pvalue)),]

ver<- rbind(unlist(t(selectCov2)[1,]),unlist(t(selectCov2)[3,]))
ver<-as.data.frame(ver)

#Exclusoes
modelo<-
coxph(Surv(ini,fim,status)~Periodo+HAART+Regime,data=PrimMudanca
,x=T)
modelo1<-
coxph(Surv(ini,fim,status)~Periodo+Regime,data=PrimMudanca,x=T)
modelo2<-
coxph(Surv(ini,fim,status)~Periodo+HAART,data=PrimMudanca,x=T)
anova(modelo,modelo1,test = "Chisq")
anova(modelo,modelo2,test = "Chisq")

```



```

#Seleccção co variaveis
deviance3 <- function(x) {
  modelo1 <- coxph(Surv(ini,fim,status)~Periodo+HAART+Idade,data=PrimMudanca,
x=T)
  modelo2 <- coxph(Surv(ini,fim,status)~Periodo+HAART+Idade+PrimMudanca[,x],d
ata=PrimMudanca,x=T)
  teste <- anova(modelo1, modelo2,test = "Chisq")
  df <- teste$Df[2]
  p <- teste[2, 4]
  resp <- data.frame(df, p)
  names(resp) <- c("Df", "Pvalue")
  resp
}

mB3<-seq(8,20)[-c(3,4)]
selectCov3<-data.frame(Variavel = names(PrimMudanca[mB3]),
t(sapply(mB3, deviance3)))
selectCov3[order(unlist(selectCov3$Pvalue)),]

```

Modelo Final (Modelo de Cox)

```

> modeloFinal<-
coxph(Surv(ini,fim,status)~Periodo+Regime+HAART,data=PrimMudanca
,x=T)
> summary(modeloFinal)
Call:
coxph(formula = Surv(ini, fim, status) ~ Periodo + Regime +
HAART,
      data = PrimMudanca, x = T)

n= 185, number of events= 94

```

	coef	exp(coef)	se(coef)	z	Pr(> z)
Periodo+2005	-0.1155	0.8909	0.4656	-0.248	0.8040
Periodo2000-2005	-0.5455	0.5796	0.3422	-1.594	0.1109
Regime2INTR+IP	0.5747	1.7766	0.2350	2.446	0.0145 *
RegimeOutro	0.9506	2.5873	0.4484	2.120	0.0340 *
HAARTNaoHAART	0.4189	1.5203	0.2764	1.516	0.1296

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Periodo+2005	0.8909	1.1225	0.3577	2.219
Periodo2000-2005	0.5796	1.7254	0.2964	1.133
Regime2INTR+IP	1.7766	0.5629	1.1209	2.816
RegimeOutro	2.5873	0.3865	1.0745	6.230

HAARTNaóHAART	1.5203	0.6578	0.8844	2.613
---------------	--------	--------	--------	-------

```

Concordance= 0.601 (se = 0.031 )
Rsquare= 0.075 (max possible= 0.974 )
Likelihood ratio test= 14.35 on 5 df, p=0.01354
Wald test = 14.13 on 5 df, p=0.01482
Score (logrank) test = 14.59 on 5 df, p=0.01229

```

Diagnóstico

Resíduos de Schoenfeld (Figura 9)

```

par(mfrow=c(2,2))
res.sch<-cox.zph(modeloFinal,transform="identity")

for(i in 1:4){
  plot(res.sch[i])
  abline(h=modeloFinal$coefficients[i],lty=4)
}

i<-5
plot(res.sch[i])
abline(h=modeloFinal$coefficients[i],lty=4)

```

Correlação linear dos resíduos das covariáveis com o tempo de sobrevivência

```
cox.zph(modeloFinal)
```

Resíduos martingala e desvios residuais (Figura 10)

```

res.mart<-resid(modeloFinal,type="martingale")
plot(res.mart,xlab="Índice",ylab="Resíduos martingala")
abline(h=0,lty=3)

resid.dev<-resid(modeloFinal,type="deviance")
plot(resid.dev,xlab="Índice",ylab="Desvios residuais")
abline(h=0,lty=3)

plot(resid.dev,predict(modeloFinal),xlab="Valores
preditos",ylab="Desvios residuais")
abline(h=0,lty=3)

qqnorm(resid.dev,xlab="Quantis amostrais",main=" ")
qqline(resid.dev,lty=3)
qqnorm(resid.dev,xlab="Quantis teóricos",ylab="Quantis
amostrais",main=" ")
qqline(resid.dev,lty=3)

```

Resíduos score (Figura 11)

```
res.sco<-resid(modeloFinal,type="dfbetas")
plot(PrimMudanca$Periodo,res.sco[,1],xlab="Período",ylab="Resíduos score")
plot(PrimMudanca$Regime,res.sco[,2],xlab="Regime",ylab="Resíduos score")
plot(PrimMudanca$HAART,res.sco[,3],xlab="HAART",ylab="Resíduos score")
```

Seleção de covariáveis (Modelo PWP)

```
deviancePWP1 <- function(x) {
  modelo <- coxph(Surv(ini,fim,status)~Periodo+PWP1[,x]+cluster(idPARTICIPANTE)+
    strata(Estrato),data=PWP1)
  df<-length(modelo$coefficients)#graus de liberdade=nr de coeficientes
  wald<-modelo$wald
  robust<-modelo$rscore
  pwald<-1-pchisq(wald,df)
  probust<-1-pchisq(robust,df)
  resp <- data.frame(df,wald, pwald, robust, probust)
  names(resp) <- c("Df", "Wald", "PWald", "Robust","PRobust")
  resp
}

varPWP<-seq(8,20)
selectCovPWP<-data.frame(Variavel = names(PWP1[varPWP]),
  t(sapply(varPWP, deviancePWP1)))
selectCovPWP[order(unlist(selectCovPWP$PWald)),]

#Seleção co variaveis 2
deviancePWP2 <- function(x) {
  if(class(PWP1[,x])!="numeric")
    nZeros<-length(levels(PWP1[,x]))-1
  else
    nZeros<-1
  teste<-as.numeric(c(coef(modeloCD4),rep(0,nZeros)))
  modelo <- coxph(Surv(ini,fim,status)~Periodo+CD4+PWP1[,x]+cluster(idPARTICIPANTE)+
    strata(Estrato),data=PWP1,init=teste)
  df<-length(modelo$coefficients)#graus de liberdade=nr de coeficientes
  wald<-modelo$wald
  robust<-modelo$rscore
  pwald<-1-pchisq(wald,df)
  probust<-1-pchisq(robust,df)
```

```

resp <- data.frame(df,wald, pwald, robust, probust)
names(resp) <- c("Df", "Wald", "PWald", "Robust","PRobust")
resp
}

varPWP2<-seq(8,20)[-c(10,11)]
selectCovPWP2<-data.frame(Variavel = names(PWP1[varPWP2]),
t(sapply(varPWP2, deviancePWP2)))
selectCovPWP2[order(unlist(selectCovPWP2$PWald)),]

```

Modelo Final (Modelo PWP)

```

> modeloCD4 <-
coxph(Surv(ini,fim,status)~Periodo+CD4+cluster(idPARTICIPANTE)+
+ strata(Estrato),data=PWP1)
> summary(modeloCD4)
Call:
coxph(formula = Surv(ini, fim, status) ~ Periodo + CD4 +
cluster(idPARTICIPANTE) + strata(Estrato), data = PWP1)
n= 357, number of events= 189

```

	coef	exp(coef)	se(coef)	robust se	z	Pr(> z)
Periodo+2005	-0.1986069	0.8198721	0.3344208	0.3187218	-0.623	0.53320
Periodo2000-2005	-0.3342343	0.7158860	0.2048858	0.2172519	-1.538	0.12394
CD4	-0.0009543	0.9990461	0.0004108	0.0003540	-2.696	0.00702 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
Periodo+2005	0.8199	1.220	0.4390	1.5312
Periodo2000-2005	0.7159	1.397	0.4676	1.0959
CD4	0.9990	1.001	0.9984	0.9997

```

Concordance= 0.568 (se = 0.038 )
Rsquare= 0.021 (max possible= 0.961 )
Likelihood ratio test= 7.46 on 3 df, p=0.05871
Wald test = 8.31 on 3 df, p=0.04002
Score (logrank) test = 7.05 on 3 df, p=0.07047, Robust =
7.64 p=0.05419

```

(Note: the likelihood ratio and score tests assume independence of observations within a cluster, the Wald and robust score tests do not).

```

plot(survfit(modeloCD4),fun="cumhaz",lty=1:4,xlab="Meses",ylab=
"Risco acumulado",mark.time=F)

```

```

legend(x="topleft",legend=c("1ª mudança", "2ª mudança", "3ª
mudança", "4ª mudança"),lty=c(1:4),bty="n")

```